Using GMM when testing for a unit root in panels where the time-series dimension is fixed

Madsen, Edith

Publication date:
2003

Document version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Madsen, E. (2003). *Using GMM when testing for a unit root in panels where the time-series dimension is fixed*. Department of Economics, University of Copenhagen.

CAM

Centre for Applied Microeconometrics

Institute of Economics University of Copenhagen

http://www.econ.ku.dk/CAM/

Using GMM when testing for a unit root in panels where the time-series dimension is fixed

Edith Madsen

2003-11

The activities of CAM are financed by a grant from The Danish National Research Foundation

Using GMM when testing for a unit root in panels where the time-series dimension is fixed

Edith Madsen[†]

Institute of Economics, University of Copenhagen, Denmark

August 2003

Abstract

In this paper we investigate GMM-based unit root inference in an autoregressive panel data model with individual-specific levels. We consider tests based on GMM estimators of the AR parameter and moment condition tests. The limiting distributions of the corresponding test statistics are derived when the AR parameter is unity and local-to-unity. This provides information about which statistics lead to valid test procedures. The performance of the valid tests in terms of their local power can then be compared. The results show that the GMM estimator of the AR parameter based on the Arellano-Bover type moment conditions, expressing that lagged differences are used as instruments for the equations in levels, can be used to detect a unit root. On the other hand, the widely used GMM estimator of the AR parameter based on the Arellano-Bond type moment conditions, expressing that lagged levels are used as instruments for the equations in first-differences, can not be used for this purpose. Instead a moment condition test of the hypothesis that the Arellano-Bond type moment conditions do not identify the AR parameter is valid as a unit root test. Finally, a simulation study demonstrates that the local power of the tests provides good approximations of their actual power in finite samples.

Keywords: Dynamic panel data model; Unit roots; GMM estimation; Local alternatives; Weak instruments

JEL classification: C12; C23

[†] Institute of Economics, University of Copenhagen, Studiestræde 6, 1455 Kbh. K., Denmark
Email: edith.madsen@econ.ku.dk

1 Introduction

In this paper we investigate unit root inference in panel data based on a generalized method of moments (GMM) approach. We consider traditional micro-panels where the cross-section dimension is much larger than the time-series dimension. There are two areas within panel data econometrics that are closely related to the subject of this paper. One is GMM estimation in dynamic micro-panels and the other is unit root inference in panel data models. The first topic has been explored for some time while the second topic is a relatively new research area that has evolved since the beginning of the 1990s. Contrary to the previous research on dynamic panels, the major part of the contributions to this new research area considers a different type of panel data where the cross-section and time-series dimensions are similar in magnitude. Reviews of the literature on GMM estimation in dynamic micro-panels are found in Baltagi (1995) and Arellano (2003) and reviews of the literature on unit root inference in panels are found in Banerjee (1999) and Baltagi & Kao (2000) of which the latter also contains a review of the first mentioned literature.

The framework considered in this paper is a first-order autoregressive panel data model with individual-specific levels. This means that we are testing the null hypothesis of each time-series process being a random walk without drift against the alternative hypothesis of each time-series process being stationary with the same autoregressive parameter for all cross-section units. In particular, the model does not allow for individual-specific linear time trends. We consider two types of GMM-based unit root tests. One is a test based on GMM estimators of the autoregressive parameter and the other is a moment condition test. The main contribution of the paper is to provide analytical results about the performance of these tests in terms of power. This is done by deriving asymptotic representations of the corresponding test statistics under both the unit root hypothesis and local-to-unity alternatives. These results are used to compare the different tests with respect to local power and hence they offer concrete information about which tests have the highest power. Furthermore, the asymptotic representations of the test statistics reveal how the nuisance parameters of the data generating process (DGP) affect the statistics. Therefore they are also very useful in relation to simulation studies as the outcomes of these are likely to depend on the particular choice of nuisance parameters in the DGP.

So far there are only few contributions to the topic of GMM-based unit root inference in micro-panels. Breitung (1997) is the first to investigate the topic. This is done within a framework which is more general than the one used in the present paper, namely an autoregressive panel data model with for individual-specific linear time trends. The performance of the GMM-based unit root tests being considered is investigated in a simulation study. Bond, Nauges & Windmeijer (2002) compare different types of unit root tests in a simulation study including GMM-based unit root tests. In a recent paper by Kruiniger (2002), which I was not aware of when I began deriving the results presented in the following, the objective seems to be similar to mine. However, our results and how they are interpreted seem to differ. So I believe the present paper is the first to provide analytical results about the power properties of GMM-based unit root tests.

In an autoregressive panel data model with individual-specific levels a widely used estimator of the autoregressive parameter is the linear GMM estimator which is obtained by taking first-differences of the equation to eliminate the individual-specific levels and by using lagged levels of the variable as instruments, see Anderson & Hsiao (1981), Holtz-Eakin, Newey & Rosen (1988) and Arellano & Bond (1991). It is well-known that the underlying moment conditions do not identify the autoregressive parameter when its value is unity. This is already noted in the earliest papers dealing with unit root inference in micro-panels, i.e. Breitung & Meyer (1994), Breitung (1997) and Harris & Tzavalis (1999). It is also well-known that the performance of this GMM estimator is poor when the autoregressive parameter is high but less than unity, see e.g. Blundell & Bond (1998) and Blundell, Bond & Windmeijer (2000), and the two things are of course related to each other. As a response to this problem Arellano & Bover (1995), suggest using additional linear moment conditions that are valid when certain restrictions are imposed on the initial values. These restrictions are satisfied when the initial values are such that the time-series processes become mean stationary. The moment conditions express that lagged first-differences of the variable can be used as instruments for the equations in levels. Arellano & Bover (1995) note that these moment conditions also identify the autoregressive parameter when its value is unity. In fact it is explicitly suggested to use the GMM estimator of the autoregressive parameter based on these moment conditions to test for a unit root. As mentioned above Breitung (1997) is the first to pursue this idea.

In this paper we derive the limiting distribution of the Arellano-Bover GMM estimator (lagged differences are used as instruments for the equations in levels) and the Arellano-Bond GMM estimator (lagged levels are used as instruments for the equations in differences) when the autoregressive parameter is unity and local-to-unity. The results show that test statistics based on the Arellano-Bover GMM estimator can be used to test for a unit root. The performance of these tests in terms of local power is quite good. This finding differs from the simulation results reported in Breitung (1997). According to these the empirical power of this type of test is quite low even for large sample sizes. The difference between our findings is probably explained by the fact that the model considered by Breitung (1997) contains individual-specific linear time trends such that second-differences instead of first-differences are used as instruments for the equations in levels. Returning to the results in this paper, they show that the Arellano-Bond GMM estimator is inconsistent and has a non-standard limiting distribution which contains nuisance parameters also in the unit root case. This means that this statistic can not be used as a unit root test. Not surprisingly, the limiting distribution of the Arellano-Bond GMM estimator is similar to the one obtained by Breitung (1997) in the unit root case. It is also interesting to note that it is similar to the one obtained by Staiger & Stock (1997) as the result of weak instruments. So as indicated above the two things are closely related. Instead, the moment condition test of the hypothesis that the Arellano-Bond type moment conditions do not identify the autoregressive parameter is a valid unit root test. This test exploits that the instruments and the endogenous variable are uncorrelated when the autoregressive parameter equals unity. It has been suggested by Bond, Nauges & Windmeijer

(2002). Our results show that the local power of this unit root test is very high under the covariance stationary alternative even for values of the autoregressive parameter very close to unity. In particular, this test performs much better than the test based on the Arellano-Bover GMM estimator under the covariance stationary alternative. A simulation study demonstrates that the local power of the tests provides a good approximation of their actual power in finite samples.

The rest of the paper is organized as follows. In Section 2, the basic model and the underlying assumptions are specified. In Section 3, the different GMM-based tests are investigated and compared by deriving the limiting distributions of the corresponding test statistics when the autoregressive parameter is unity and local-to-unity. We consider tests based on either the Arellano-Bover or the Arellano-Bond type moment conditions. In Section 4, the analytical results are illustrated in a simulation study. Section 5 provides some concluding remarks.

2 The model and assumptions

We consider the following first-order autoregressive panel data model

$$y_{it} = \rho y_{it-1} + (1-\rho)\alpha_i + \varepsilon_{it} \quad \text{for } i = 1, ..., N \text{ and } t = 1, ..., T \tag{1}$$

where $-1 < \rho \leq 1$ and for every $i = 1, ..., N$ the sequence $\{\varepsilon_{it}\}_{t=1}^{\infty}$ is white noise. For notational convenience we assume that the initial values y_{i0} are observed such that the actual number of observations over time equals $T+1$. When $|\rho| < 1$ the time-series process for y_{it} defined by equation (1) is stationary. The precise meaning of this statement is usually that certain restrictions can be imposed on the initial value y_{i0} such that the time-series process becomes covariance stationary. On the other hand, when $\rho = 1$ the time-series process for y_{it} is a random walk which is obviously not stationary in the sense described above. So using the standard time-series terminology, the value of the AR parameter can be used to distinguish between stationarity and a particular form of non-stationarity of the time-series processes. Another way of interpreting the value of the AR parameter, which also takes the cross-section dimension into account, is as describing where the persistency in the time series $\{y_{it}\}_{t=0}^{T}$ for $i = 1, ..., N$ comes from. According to the model in (1) there are two sources of persistency. One is the autoregressive mechanism described by the AR parameter which is the same for all individuals. The other is the unobserved individual-specific effects described by the term α_i. Everything else equal, a higher AR parameter means that more persistency is attributed to the common autoregressive mechanism and less to the unobserved individual-specific effects. An AR parameter of unity can then be considered as an extreme case where all persistency in the time series is attributed to the autoregressive mechanism.

To specify the model in (1) further, the following general assumptions are imposed.

Assumption 1 (***Standard assumptions***)

(i) *ε_{it} is independent across i, t with $E(\varepsilon_{it}) = 0$ and $E(\varepsilon_{it}^2) = \sigma_{i\varepsilon}^2$*

(ii) *α_i is iid across i and independent of $\varepsilon_{i1}, ..., \varepsilon_{iT}$ with $E(\alpha_i) = 0$ and $E(\alpha_i^2) = \sigma_\alpha^2$*

(iii) ε_{it} *is independent of* y_{i0} *for* $t = 1, ..., T$

(i) states that the errors ε_{it} are independent over both cross-section units and time and only allowed to be heteroskedastic over cross-section units not over time, (ii) is the standard assumption concerning the unobserved individual-specific effects, and (iii) states that the initial values of the time-series processes are independent of all subsequent innovations. The assumptions are stronger than the ones often encountered in the literature. Often the errors ε_{it} are assumed to be serially uncorrelated and allowed to be heteroskedastic over time, see for example Section 8 in Baltagi (1995) and Section 6 in Arellano (2003), whereas according to (i) the errors ε_{it} are assumed to be independent and homoskedastic over time. The stronger assumptions are imposed in order to simplify the presentation of the asymptotic properties of the statistics considered in Section 3.

The initial values y_{i0} are specified as follows.

Assumption 2 (***Initial values***)
For $-1 < \rho \leq 1$ *the initial values can be decomposed as* $y_{i0} = \mathbf{1}_{\{|\rho|<1\}}\alpha_i + \sqrt{\tau(\rho)}\varepsilon_{i0}$ *where* ε_{i0} *is independent across i and independent of* α_i *with* $E(\varepsilon_{i0}) = 0$ *and* $E(\varepsilon_{i0}^2) = \sigma_{i\varepsilon}^2$. *The scaling function* $\tau(\rho)$ *can be on the following forms: (i)* $\tau(\rho) = \tau$ *where* $0 \leq \tau < \infty$ *when* $-1 < \rho \leq 1$, *(ii)* $\tau(\rho) = 1/(1-\rho^2)$ *when* $|\rho| < 1$.

When $\rho = 1$ the assumption states that the variance of the initial value y_{i0} equals $\tau\sigma_{i\varepsilon}^2$. This assumption is slightly stronger compared to letting the initial values be completely heteroskedastic across units. It is stronger because it restricts the variance of the initial value to be proportional to the variance of the errors ε_{it} with a proportionality factor τ which is the same for all units. The assumption is imposed in order to simplify the derivations of the asymptotic properties of the statistics considered in Section 3. When $|\rho| < 1$ the time-series processes for y_{it} are always assumed to be mean stationary, and we consider two different situations characterized by the variance of the initial deviation from the stationary level $(y_{i0} - \alpha_i)$. In (i) the variance of $(y_{i0} - \alpha_i)$ equals $\tau\sigma_{i\varepsilon}^2$ whereas in (ii) it equals $\sigma_{i\varepsilon}^2/\left(1-\rho^2\right)$. In the latter case, the time-series processes are covariance stationary. The distinction between mean stationarity and covariance stationarity appears to be important for the results in Section 3 where it will be discussed further. Also note that this specification of the initial values together with Assumption 1 implies that ε_{i0} is independent of ε_{it} for $t = 1, ..., T$. For a discussion of the assumptions concerning the initial values in dynamic panel data models see for example Section 6.4 in Arellano (2003).

Finally, we need some technical assumptions in order to derive the asymptotic properties of the statistics specified in Section 3 by applying standard asymptotic theory, see for example White (2001).

Assumption 3 (***Technical assumptions***)

(i) $E\left(\alpha_i^4\right) < \infty$

(ii) $E|\varepsilon_{it}|^{4+\delta} < K < \infty$ *for some* $\delta > 0$ *and all* $i = 1, ..., N, t = 0, ..., T$

(iii) $\frac{1}{N}\sum_{i=1}^{N}\sigma_{i\varepsilon}^2 \to \sigma_{2\varepsilon} > 0$ *as* $N \to \infty$

(iv) $\frac{1}{N}\sum_{i=1}^{N}\sigma_{i\varepsilon}^{4} \to \sigma_{4\varepsilon}$ *as* $N \to \infty$

(i) states that the individual-specific terms α_i have finite moments of fourth order, (ii) states that the errors ε_{it} have uniformly bounded moments of order slightly greater than four, and (iii) and (iv) state that the cross-section averages of the variances and the squared variances converge to well-defined limits as the cross-section dimension tends to infinity.

Before we turn to the test statistics a short comment on the specification of the autoregressive panel data model in (1). Usually the following model is considered

$$y_{it} = \rho y_{it-1} + \eta_i + \varepsilon_{it} \tag{2}$$

see for example Section 8 in Baltagi (1995) and Section 6 in Arellano (2003). When $|\rho| < 1$ this is just a re-parametrization of the model in (1). On the other hand, when $\rho = 1$ the two models lead to different time-series processes for y_{it} as the process defined by (2) in this case is a random walk with drift. Therefore, with respect to interpretation, the unit root hypothesis within the model defined by (1) is preferable to the unit root hypothesis within the model defined by (2). If there is any prior belief that the variable of interest contains an individual-specific linear time trend, the model should be formulated such that this is allowed for under both the null hypothesis and the alternative hypothesis, see Breitung (1997). Unit root inference within an autoregressive panel data model with individual-specific linear time trends will not be investigated in this paper. However, the difference between the two specifications of the autoregressive panel data model will be discussed further in Section 3.

3 The GMM-based test statistics

The testing problem is given by the null hypothesis H_0 and the alternative hypothesis H_A which are

$$H_0 : \rho = 1 \qquad H_A : |\rho| < 1 \tag{3}$$

In the following we consider local alternatives where ρ is modelled as being local-to-unity, i.e.

$$\rho = 1 - \frac{c}{\sqrt{N}} \quad \text{where } c > 0 \tag{4}$$

This means that we consider $c = (1-\rho)\sqrt{N}$ as being a constant such that the AR parameter ρ is in a $1/\sqrt{N}$ neighborhood of unity. The idea is that the asymptotic representations of statistics derived under the local-to-unity sequence for ρ will provide good approximations to the actual behavior of the statistics for values of the AR parameter close to unity.

The assumption concerning the initial values y_{i0} is important as y_{i0} affects the variables y_{it} for $t = 1, ..., T$. In particular, under the local-to-unity sequence for AR parameter ρ given by (4) it matters how this affects the initial values. According to Assumption 2, the variance of $\left(y_{i0} - \mathbf{1}_{\{|\rho|<1\}}\alpha_i\right)$ equals $\tau(\rho)\,\sigma_{i\varepsilon}^2$ where $\tau(\rho)$ can be on the following two forms

$$(i) \text{ and } -1 < \rho \leq 1 : \qquad \tau(\rho) = \tau < \infty \tag{5}$$

$$(ii) \text{ and } -1 < \rho < 1 : \qquad \tau(\rho) = \frac{1}{1-\rho^2} \tag{6}$$

Under the local-to-unity sequence for ρ given in (4) this corresponds to

$$(i) \text{ and } c \geq 0: \qquad \tau(\rho) = \tau = O(1) \tag{7}$$

$$(ii) \text{ and } c > 0: \qquad \tau(\rho) = \frac{\sqrt{N}}{2c} + o\left(N^{-\frac{1}{2}}\right) = O\left(N^{\frac{1}{2}}\right) \tag{8}$$

where (8) follows by Lemma 1 in Appendix A.1. In (ii) the behavior of the variable y_{it} as N tends to infinity is dominated by the initial deviation from the stationary level $(y_{i0} - \alpha_i)$ as the variability of this term is of order $\sqrt{N}$ whereas the variability of the remaining terms in y_{it} is bounded. In (i) the variability of all terms in y_{it} is bounded. This means that the asymptotic behavior of y_{it} is similar under the mean stationary local alternative and under the null hypothesis of ρ being unity but differs under the covariance stationary local alternative.

3.1 The Arellano-Bover type moment conditions

The equation in (1) can be rewritten as the following regression model

$$\begin{aligned} y_{it} &= \rho y_{it-1} + v_{it} \\ v_{it} &= (1-\rho)\alpha_i + \varepsilon_{it} \end{aligned} \qquad \text{for } i = 1, ..., N \text{ and } t = 2, ..., T \tag{9}$$

As suggested by Arellano & Bover (1995) it is possible to use lagged first-differences as instruments for the equations in levels when a certain restriction is imposed on the initial values. The first-differences Δy_{it} can be expressed in terms of Δy_{i1} and the errors $\Delta\varepsilon_{i2}, ..., \Delta\varepsilon_{it}$ as follows

$$\Delta y_{it} = \rho^{t-1}\Delta y_{i1} + \rho^{t-2}\Delta\varepsilon_{i2} + ... + \Delta\varepsilon_{it} \quad \text{for } t = 2, ..., T \tag{10}$$

Using this together with $\Delta y_{i1} = (\rho - 1)(y_{i0} - \alpha_i) + \varepsilon_{i1}$ we see that when $\rho = 1$ then Δy_{it-s} and $v_{it} = (1-\rho)\alpha_i + \varepsilon_{it}$ are always uncorrelated for $s = 1, ...t - 1$. When $|\rho| < 1$ this is true if and only if α_i and Δy_{i1} are uncorrelated, i.e.

$$E(\alpha_i \Delta y_{i1}) = 0 \quad \text{for } i = 1, ..., N \tag{11}$$

To see that this only concerns the initial value we use the expression for Δy_{i1} together with $E(\alpha_i \varepsilon_{1i}) = 0$. Hence, the restriction in (11) is equivalent to $(\rho - 1) E(\alpha_i (y_{i0} - \alpha_i)) = 0$, which expresses that the initial deviation from the stationary level is uncorrelated with the stationary level itself. This holds when the initial values are mean stationary and hence the restriction in (11) is satisfied under Assumption 2. Therefore, under Assumption 1 and 2 the following $m = \frac{1}{2}T(T-1)$ moment conditions which are linear in the AR parameter ρ holds

$$E(\Delta y_{it-s} v_{it}) = 0 \quad \text{for } t = 2, ..., T \text{ and } s = 1, ..., t - 1 \tag{12}$$

Using stacked notation they can be expressed as

$$E(Z_{i1}'(y_i - \rho y_{i,-1})) = 0 \tag{13}$$

where y_i and $y_{i,-1}$ are $(T-1)\times 1$ vectors defined as $y_i = (y_{i2}, ..., y_{iT})'$ and $y_{i,-1} = (y_{i1}, ..., y_{iT-1})'$ and Z_{i1} is the $(T-1)\times m$ matrix defined as

$$Z_{i1} = \begin{bmatrix} \Delta y_{i1} & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \Delta y_{i1} & \Delta y_{i2} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & \Delta y_{i1} & \cdots & \Delta y_{iT-1} \end{bmatrix} \tag{14}$$

As first noted by Arellano & Bover (1995), these moment conditions also identify the AR parameter ρ when its true value is unity.

The GMM estimator based on the moment conditions in (13) is then defined as the value of ρ which minimizes

$$Q_N(\rho) = f_N(\rho)' W_{1N} f_N(\rho) \tag{15}$$

where $f_N(\rho) = \frac{1}{N}\sum_{i=1}^{N} Z_{i1}'(y_i - \rho y_{i,-1})$ and W_{1N} is a stochastic positive definite weighting matrix. The GMM estimator of ρ is then given by the following expression

$$\hat{\rho}_I = \left(\sum_{i=1}^{N} y_{i,-1}' Z_{i1} W_{1N} \sum_{i=1}^{N} Z_{i1}' y_{i,-1}\right)^{-1} \left(\sum_{i=1}^{N} y_{i,-1}' Z_{i1} W_{1N} \sum_{i=1}^{N} Z_{i1}' y_i\right) \tag{16}$$

The estimator $\hat{\rho}_I$ is consistent (as $N \to \infty$ and T fixed) for all positive definite $O_P(1)$ weighting matrices but differ in terms of asymptotic variance. The weighting matrix resulting in an asymptotically efficient estimator within the class of GMM estimators defined in (16) is given by

$$W_{1N} = \left(\frac{1}{N}\sum_{i=1}^{N} Z_{i1}'\hat{v}_i\hat{v}_i' Z_{i1}\right)^{-1} \tag{17}$$

where $\hat{v}_i$ are residuals from an initial consistent estimator of ρ. We refer to this as the optimal two-step Arellano-Bover GMM estimator. In the following we will use the standard IV estimator as a consistent one-step estimator, i.e. we use the weighting matrix $W_{1N} = \left(\frac{1}{N}\sum_{i=1}^{N} Z_{i1}' Z_{i1}\right)^{-1}$ in the first step. Unless $(1-\rho)^2 \sigma_\alpha^2 = 0$, there is no one-step estimator which is asymptotically equivalent to the two-step estimator even when the errors ε_{it} are homoskedastic across units. As $(1-\rho)^2 \sigma_\alpha^2 = 0$ when $\rho = 1$ the two estimators are asymptotically equivalent in this case even when the errors ε_{it} are heteroskedastic across units. In fact the two estimators are asymptotically equivalent under the local-to-unity sequence for ρ defined in (4), for details see Appendix A.2.

The limiting distribution of the optimal two-step Arellano-Bover GMM estimator $\hat{\rho}_I$ under both the null hypothesis when ρ is unity and the local alternative when ρ is local-to-unity is provided in Proposition 1 below. Under the local alternative we distinguish between the two situations where the time-series processes for y_{it} are respectively mean stationary and covariance stationary, as expressed by (i) and (ii) in Assumption 2.

Proposition 1 *Under Assumption 1, 2, 3 and the local-to-unity sequence for ρ given by $\rho = 1 - c/\sqrt{N}$ for $c \geq 0$, the limiting distribution of the optimal two-step Arellano-Bover GMM estimator $\hat{\rho}_I$ is given* *by*

$$(i) \text{ and } c \geq 0: \qquad \sqrt{N}\left(\hat{\rho}_I - \rho\right) \xrightarrow{w} N\left(0, \frac{\sigma_{4\varepsilon}}{\sigma_{2\varepsilon}^2}\frac{2}{T\left(T-1\right)}\right) \quad \text{as } N \to \infty \tag{18}$$

$$(ii) \text{ and } c > 0: \qquad \sqrt{N}\left(\hat{\rho}_I - \rho\right) \xrightarrow{w} N\left(0, 4\frac{\sigma_{4\varepsilon}}{\sigma_{2\varepsilon}^2}\frac{2}{T\left(T-1\right)}\right) \quad \text{as } N \to \infty \tag{19}$$

The proof of Proposition 1 is given in Appendix A.2. The proposition shows that under both the null hypothesis and the local alternative, the Arellano-Bover GMM estimator $\hat{\rho}_I$ is $\sqrt{N}$-consistent and its limiting variance depends on T and $\sigma_{2\varepsilon}^2/\sigma_{4\varepsilon}$ and is decreasing in both. Further, the limiting variance of $\hat{\rho}_I$ is 4 times bigger under the covariance stationary alternative than under the mean stationary alternative. This is because the covariance between the endogenous variable $y_{i,-1}$ and the instrument Z_{i1} in the covariance stationary case is half of that in the mean stationary case. Also, according to the proposition the parameter σ_α^2 does not appear in the limiting distribution of $\hat{\rho}_I$ under the local alternative. First of all, this is because the term α_i does not appear in the instruments under the assumption about mean stationarity which is always imposed. This implies that the partial correlation coefficient between the endogenous variable $y_{i,-1}$ and the instrument Z_{i1} is not affected by α_i. In addition, the regression error v_{it} equals $c\alpha_i/\sqrt{N} + \varepsilon_{it}$ under the local-to-unity sequence for ρ, such that asymptotically as N tends to infinity, the behavior of v_{it} is dominated by ε_{it}. This indicates that the asymptotic representations in Proposition 1 are only appropriate when σ_α^2 is of order less than N. If this is not the case, the distributions in Proposition 1 are not expected to provide good approximations to the actual distribution of $\hat{\rho}_I$.

The usual t-statistic corresponding to the unit root hypothesis is given by the following expression

$$t_I = \left(\sum_{i=1}^{N} y_{i,-1}' Z_{i1}\left(\sum_{i=1}^{N} Z_{i1}'\hat{v}_i\hat{v}_i' Z_{i1}\right)^{-1}\sum_{i=1}^{N} Z_{i1}' y_{i,-1}\right)^{\frac{1}{2}}\left(\hat{\rho}_I - 1\right) \tag{20}$$

When the errors ε_{it} are homoskedastic across units such that $\sigma_{4\varepsilon} = \sigma_{2\varepsilon}^2$, the limiting variance of $\hat{\rho}_I$ only depends on T. In this case it is possible to use a normalized coefficient statistic when testing the unit root hypothesis. The statistic is defined as

$$\bar{t}_I = \sqrt{\frac{T\left(T-1\right)}{2}}\sqrt{N}\left(\hat{\rho}_I - 1\right) \tag{21}$$

The limiting distributions of these test statistics are provided in Proposition 2 below.

Proposition 2 *Under Assumption 1, 2, 3 and the local-to-unity sequence for ρ given by $\rho = 1 - c/\sqrt{N}$ for $c \geq 0$, the limiting distribution of the t-statistic t_I is given by*

$$(i) \text{ and } c \geq 0: \qquad t_I \xrightarrow{w} N\left(-c\sqrt{\frac{\sigma_{2\varepsilon}^2}{\sigma_{4\varepsilon}}\frac{T\left(T-1\right)}{2}}, 1\right) \quad \text{as } N \to \infty \tag{22}$$

$$(ii) \text{ and } c > 0: \qquad t_I \xrightarrow{w} N\left(-\frac{c}{2}\sqrt{\frac{\sigma_{2\varepsilon}^2}{\sigma_{4\varepsilon}}\frac{T\left(T-1\right)}{2}}, 1\right) \quad \text{as } N \to \infty \tag{23}$$

The limiting distribution of the normalized coefficient statistic $\bar{t}_I$ *is given by*

$$(i)\ and\ c \geq 0: \qquad \bar{t}_I \xrightarrow{w} N\left(-c\sqrt{\frac{T(T-1)}{2}}, \frac{\sigma_{4\varepsilon}}{\sigma_{2\varepsilon}^2}\right) \quad as\ N \to \infty \tag{24}$$

$$(ii)\ and\ c > 0: \qquad \bar{t}_I \xrightarrow{w} N\left(-c\sqrt{\frac{T(T-1)}{2}}, 4\frac{\sigma_{4\varepsilon}}{\sigma_{2\varepsilon}^2}\right) \quad as\ N \to \infty \tag{25}$$

The proof of Proposition 2 is given in Appendix A.2. The proposition shows that in the unit root case when $c = 0$, the limiting distribution of the t-statistic t_I is standard normal. This means that unit root inference is carried out by employing critical values from the standard normal distribution. In addition, the local power of the test based on t_I is increasing in both T and $\sigma_{2\varepsilon}^2/\sigma_{4\varepsilon}$. Also the local power of the test based on t_I is higher under the mean stationary alternative compared to the covariance stationary alternative as the location parameter is twice as large in absolute value in the first case compared to the latter case. When $\sigma_{4\varepsilon} = \sigma_{2\varepsilon}^2$ the test based on the t-statistic t_I and the test based on the normalized coefficient statistic $\bar{t}_I$ are asymptotically equivalent under the mean stationary alternative. On the other hand, under the covariance stationary alternative, the test based on the normalized coefficient $\bar{t}_I$ has higher local power than the test based on the t-statistic t_I. Letting Φ denote the distribution function of the standard normal, this follows as $P(t_I < q) = \Phi\left(q + c/2\sqrt{T(T-1)/2}\right) \leq \Phi\left(q/2 + c/2\sqrt{T(T-1)/2}\right) = P(\bar{t}_I < q)$ when $q \leq 0$ which is the case for a one-sided test at a significance level of 5%, i.e. $q \approx -1.645$. However, when $\sigma_{4\varepsilon} \neq \sigma_{2\varepsilon}^2$ the test based on the normalized coefficient statistic $\bar{t}_I$ will be distorted. In a one-sided test it will tend to reject the null hypothesis too often when $\sigma_{4\varepsilon} > \sigma_{2\varepsilon}^2$, i.e. the test is over-sized. The opposite is true when $\sigma_{4\varepsilon} < \sigma_{2\varepsilon}^2$.

The unit root test based on the t-statistic t_I is asymptotically equivalent to the unit root test suggested by Breitung & Meyer (1994) which is based on a t-statistic corresponding to a least squares regression of $(y_{it} - y_{i0})$ on $(y_{it-1} - y_{i0})$, see the results in Madsen (2003). The advantage of using the Breitung-Meyer unit root test is that it is invariant with respect to the individual-specific levels even in finite samples. In particular, its power does not depend on the individual-specific terms α_i. This is not the case for a test based on the Arellano-Bover GMM estimator $\hat{\rho}_I$.

Finally, as suggested by Breitung (1997) we consider a unit root test which is a test of the validity of certain moment conditions. The test is similar to the test for validity of the instruments suggested by Hansen (1982). The hypothesis is expressed as the following $m = \frac{1}{2}T(T-1)$ moment conditions

$$E(Z_{i1}'\Delta y_i) = 0 \tag{26}$$

which are valid when $\rho = 1$ but not when $|\rho| < 1$. The moment conditions can be tested using the statistic

$$J_I = \sum_{i=1}^{N} \Delta y_i' Z_{i1} \left(\sum_{i=1}^{N} Z_{i1}' \Delta y_i \Delta y_i' Z_{i1}\right)^{-1} \sum_{i=1}^{N} Z_{i1}' \Delta y_i \tag{27}$$

Proposition 3 *Under Assumption 1, 2, 3 and local-to-unity sequence for* ρ *given by* $\rho = 1 - c/\sqrt{N}$ *for*

$c \geq 0$, *the limiting distribution of the statistic* J_I *is given by*

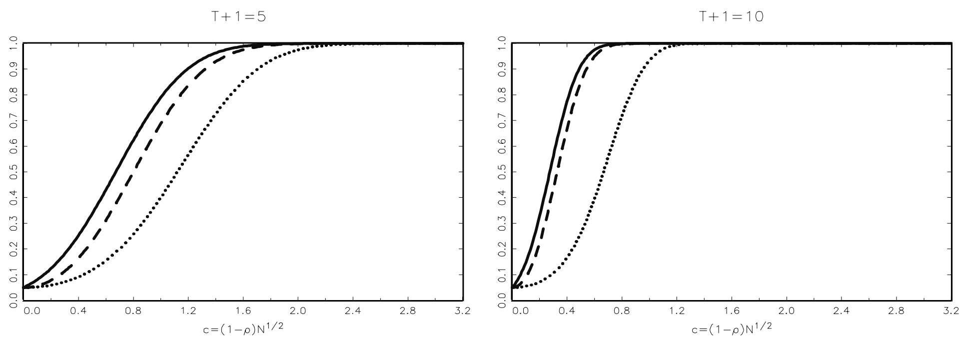
$$(i) \text{ and } c \geq 0: \qquad J_I \xrightarrow{w} \chi^2_{\frac{1}{2}T(T-1)}\left(c^2 \frac{\sigma^2_{2\varepsilon}}{\sigma_{4\varepsilon}} \frac{T(T-1)}{2}\right) \quad \text{as } N \to \infty \tag{28}$$

$$(ii) \text{ and } c > 0: \qquad J_I \xrightarrow{w} \chi^2_{\frac{1}{2}T(T-1)}\left(\frac{1}{4}c^2 \frac{\sigma^2_{2\varepsilon}}{\sigma_{4\varepsilon}} \frac{T(T-1)}{2}\right) \quad \text{as } N \to \infty \tag{29}$$

The proof of Proposition 3 is given in Appendix A.2. The proposition shows that in the unit root case, the limiting distribution of the statistic J_I is a χ^2-distribution with $\frac{1}{2}T(T-1)$ degrees of freedom. So unit root inference based on this statistic is carried out by employing critical values from this χ^2-distribution. Furthermore, under the local alternative the limiting distribution of J_I is a non-central χ^2-distribution with $\frac{1}{2}T(T-1)$ degrees of freedom and a non-centrality parameter which is a function of T and $\sigma^2_{2\varepsilon}/\sigma_{4\varepsilon}$ and is increasing in both. Similar to the results in Proposition 2, the non-centrality parameter is 4 times bigger under the mean stationary local alternative compared to the covariance stationary local alternative.

The unit root tests based on the squared t-statistic t_I^2 and J_I are asymptotically equivalent when the number of observations $T+1$ equals 3. When $T+1 > 3$ this is not the case. To investigate the difference between the tests, the local power of the tests based on t_I (one-sided), t_I^2 and J_I when $T+1$ equals 5 or 10 is shown in Figure 1. The figure shows the local power of the tests as a function of $c = (1-\rho)\sqrt{N}$ under the assumption about mean stationarity and $\sigma_{4\varepsilon} = \sigma^2_{2\varepsilon}$. We see that the local power of a unit root test based on t_I or t_I^2 is always higher than that of a test based on J_I. As noted by Breitung (1997) this is not surprising as the test based on J_I is testing $\frac{1}{2}T(T-1)$ restrictions whereas the test based on t_I is testing one restriction. Nevertheless, an advantage of the test based on J_I is that it is invariant with respect to the individual-specific levels even in finite samples. In particular, its power does not depend on the individual-specific terms α_i. As noted earlier, a test based on t_I does not have this property.

Figure 1: Local power of unit root tests based on t_I (solid line), t_I^2 (dashed line) and J_I (dotted line) under mean stationarity



3.2 The Arellano-Bond type moment conditions

Taking first-differences of all variables in (1) yields the following regression model

$$\begin{aligned} \Delta y_{it} &= \rho \Delta y_{it-1} + \Delta v_{it} \\ \Delta v_{it} &= \Delta \varepsilon_{it} \end{aligned} \qquad \text{for } i = 1, ..., N \text{ and } t = 2, ..., T \tag{30}$$

The Arellano-Bond GMM estimator is based on the following $m = \frac{1}{2}T(T-1)$ linear moment conditions

$$E(y_{it-s}\Delta v_{it}) = 0 \quad \text{for } t = 2, ..., T \text{ and } s = 2, ..., t \tag{31}$$

They are valid under Assumption 1 and unlike the Arellano-Bover type moment conditions in (12) they do not require additional restrictions on the initial values. In spite of this, Assumption 2 concerning the initial values is always imposed in the following. Using stacked notation, the moment conditions above can be expressed as

$$E(Z_{i2}'(\Delta y_i - \rho \Delta y_{i,-1})) = 0 \tag{32}$$

where Δy_i and $\Delta y_{i,-1}$ are $(T-1)\times 1$ vectors defined as $y_i = (\Delta y_{i2}, ..., \Delta y_{iT})'$ and $\Delta y_{i,-1} = (\Delta y_{i1}, ..., \Delta y_{iT-1})'$ and Z_{i2} is the $(T-1)\times m$ matrix defined as

$$Z_{i2} = \begin{bmatrix} y_{i0} & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & y_{i0} & y_{i1} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & y_{i0} & \cdots & y_{iT-2} \end{bmatrix} \tag{33}$$

When $|\rho| < 1$ the moment conditions in (32) identify the parameter ρ. When $\rho = 1$ this is not so as $E(Z_{i2}'\Delta y_{i,-1}) = 0$ in this case. Given this information, it comes as no surprise that the problem of weak instruments might appear as ρ approaches unity. The weak instrument problem is characterized by the partial correlation between the endogenous variable and the instrument being low which in Staiger & Stock (1997) is modeled as local-to-zero. More specifically, Staiger & Stock (1997) characterize the problem of weak instruments as the situation where the covariance between the endogenous variable and the instrument is $O(N^{-\frac{1}{2}})$ and the variance of the instrument is $O(1)$. Under the local-to-unity sequence for ρ given in (4) we do not necessarily have this situation. Under Assumption 2 (i) when the time-series processes are mean stationary we do have the usual weak instrument problem whereas under Assumption 2 (ii) when the time-series processes are covariance stationary we do not. In the latter case, we have a different situation where the covariance between the endogenous variable $\Delta y_{i,-1}$ and the instrument Z_{i2} is $O(1)$ while the variance of the instrument Z_{i2} is $O(N^{\frac{1}{2}})$. More specifically, the probability limits of $\frac{1}{N}\sum_{i=1}^{N}\Delta y_{i,-1}' Z_{i2}$ and $\frac{1}{N^{3/2}}\sum_{i=1}^{N} Z_{i2}'Z_{i2}$ are both well-defined and different from zero. However, the probability limit of $\frac{1}{N^{3/2}}\sum_{i=1}^{N} Z_{i2}'Z_{i2}$ is a singular matrix as the behavior of y_{it} for all $t = 0, ..., T$ is dominated by $(y_{i0} - \alpha_i)$ as N tends to infinity.

If we consider the specification of the autoregressive panel data model in (2), the mean stationary level is $\eta_i/(1-\rho)$. In this case, the behavior of y_{it} is dominated by $\eta_i/(1-\rho)$ as N tends to infinity under the local-to-unity sequence for ρ given by (4). This is true under both Assumption 2 (i) and (ii). It means that the sample covariance between the endogenous variable $\Delta y_{i,-1}$ and the instrument Z_{i2}

is $O(N^{-\frac{1}{2}})$ under Assumption 2 (i) and $O(1)$ under Assumption 2 (ii) whereas the sample variance of the instrument Z_{i2} is $O(N)$. This means that the sample moments must be normalized differently in order to converge. Also as above, the probability limit of $\frac{1}{N^2}\sum_{i=1}^{N} Z_{i2}'Z_{i2}$ is a singular non-zero matrix. Blundell & Bond (1998) and Blundell, Bond & Windmeijer (2000) consider the model in (2) but do not recognize that letting ρ approach unity is somewhat different from the problem of weak instruments as characterized in Staiger & Stock (1997).

The GMM estimator based on the Arellano-Bond type moment conditions in (32) is given by the following expression

$$\hat{\rho}_{II} = \left(\sum_{i=1}^{N} \Delta y_{i,-1}' Z_{i2} W_{2N} \sum_{i=1}^{N} Z_{i2}' \Delta y_{i,-1}\right)^{-1} \left(\sum_{i=1}^{N} \Delta y_{i,-1}' Z_{i2} W_{2N} \sum_{i=1}^{N} Z_{i2}' \Delta y_{i}\right) \tag{34}$$

In the following we consider the standard IV estimator obtained by using the weighting matrix $W_{2N} = \left(\frac{1}{N}\sum_{i=1}^{N} Z_{i2}'Z_{i2}\right)^{-1}$. In Proposition 4 below we provide the limiting distribution of this estimator under the local-to-unity sequence for ρ and the assumption about mean stationarity which according to the discussion above corresponds to the usual weak instruments problem.

Proposition 4 *Under Assumption 1, 2(i), 3 and the local-to-unity sequence for ρ given by $\rho = 1 - c/\sqrt{N}$ for $c \geq 0$, the limiting distribution of the Arellano-Bond GMM estimator $\hat{\rho}_{II}$ is given by*

$$\hat{\rho}_{II} - \rho \xrightarrow{w} \frac{X_2'\tilde{\Sigma}_{11}^{-1}X_1}{X_2'\tilde{\Sigma}_{11}^{-1}X_2} \quad \text{as } N \to \infty \tag{35}$$

where X_1 and X_2 are both $\frac{1}{2}T(T-1) \times 1$ vectors which are distributed as

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ -cq \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{bmatrix}\right) \tag{36}$$

and expressions for the vector q and the matrices $\tilde{\Sigma}_{11}$, Σ_{11}, Σ_{12} and Σ_{22} are found in Lemma 6 in Appendix A.3.

The proof of Proposition 4 is given in Appendix A.3. The proposition shows that under the local-to-unity sequence for ρ the Arellano-Bond GMM estimator $\hat{\rho}_{II}$ is not consistent. Instead $(\hat{\rho}_{II} - \rho)$ is asymptotically distributed as the ratio of quadratic forms in two jointly normal variables. The asymptotic distribution is similar to the one obtained by Breitung (1997) where it is derived under the assumption that $\rho = 1$ and $y_{i0} = 0$ for $i = 1, ..., N$. If the limiting distribution in (35) did not contain nuisance parameters under the unit root hypothesis, we could obtain empirical quantiles of this distribution by simulating it. Unit root inference could then be carried out by employing the appropriate empirical quantiles as critical values. However, in general the distribution depends on the parameters $\sigma_{2\varepsilon}^2/\sigma_{4\varepsilon}$ and τ under the unit root hypothesis. So even when the errors ε_{it} are homoskedastic across units, the nuisance parameter τ appears in the distribution. Hence, the Arellano-Bond GMM estimator $\hat{\rho}_{II}$ can not be used for unit root inference.

A more promising way to make use of the moment conditions in (32) in relation to unit root inference is by using that they do not identify ρ when its value is unity. This is the hypothesis that the endogenous variable and the instruments are uncorrelated, i.e.

$$E\left(Z_{i2}'\Delta y_{i-1}\right)=0 \tag{37}$$

The test of this hypothesis can be considered as a test for under-identification, see Arellano, Hansen & Sentana (1999). It is suggested by Bond, Nauges & Windmeijer (2002) as a unit root test. The moment conditions in (37) are tested using the statistic

$$J_{II}=\sum_{i=1}^{N}\Delta y_{i,-1}'Z_{i2}\left(\sum_{i=1}^{N}Z_{i2}'\Delta y_{i,-1}\Delta y_{i,-1}'Z_{i2}\right)^{-1}\sum_{i=1}^{N}Z_{i2}'\Delta y_{i,-1} \tag{38}$$

Proposition 5 *Under Assumption 1, 2(i), 3 and the local-to-unity sequence for ρ given by $\rho=1-c/\sqrt{N}$ for $c\geq 0$, the limiting distribution of the statistic J_{II} is given by*

$$J_{II}\xrightarrow{w}\chi^2_{\frac{1}{2}T(T-1)}\left(c^2\kappa\right)\quad \textit{as } N\to\infty \tag{39}$$

where

$$\kappa=\frac{\sigma_{2\varepsilon}^2}{\sigma_{4\varepsilon}}\left(\frac{(T-1)(T-2)}{2}+(T-1)\frac{\tau^2}{\sigma_\alpha^2\sigma_{2\varepsilon}/\sigma_{4\varepsilon}+\tau}\right) \tag{40}$$

Under Assumption 1, 2(ii), 3 and the local-to-unity sequence for ρ given by $\rho=1-\tilde{c}/N$ for $\tilde{c}>0$, the limiting distribution of the statistic J_{II} is given by

$$J_{II}\xrightarrow{w}\chi^2_{\frac{1}{2}T(T-1)}\left(\tilde{c}\frac{\sigma_{2\varepsilon}^2}{\sigma_{4\varepsilon}}\frac{T-1}{2}\right)\quad \textit{as } N\to\infty \tag{41}$$

The proof of Proposition 5 is given in Appendix A.3. The proposition shows that in the unit root case the limiting distribution of J_{II} is a χ^2-distribution with $\frac{1}{2}T(T-1)$ degrees of freedom. Under the mean stationary local alternative, the limiting distribution of J_{II} is a non-central χ^2-distribution with $\frac{1}{2}T(T-1)$ degrees of freedom and a non-centrality parameter which depends on T and all of the nuisance parameters. The local power is increasing in T and τ and decreasing in σ_α^2. Under the assumption about covariance stationarity, we consider a different local alternative where $\rho=1-\tilde{c}/N$ for $\tilde{c}>0$. In this case, the limiting distribution of J_{II} is again a non-central χ^2-distribution with $\frac{1}{2}T(T-1)$ degrees of freedom and a non-centrality parameter which depends on T and $\sigma_{2\varepsilon}^2/\sigma_{4\varepsilon}$ and is increasing in both. As explained before, the behavior of the variable y_{it} is dominated by $(y_{i0}-\alpha_i)$ under the covariance stationary local alternative. This indicates that the limiting distribution in (41) is only appropriate when the variance of $(y_{i0}-\alpha_i)$ is of higher order than the variance of all other terms in y_{it}.

To compare the unit root test based on J_{II} to the unit root tests based on t_I (one-sided), $\bar{t}_I$ (one-sided) and J_I, the local power of these tests when $T+1$ equals 5 or 10 are shown in Figure 2. The figure shows the local power of the tests as a function of $c=(1-\rho)\sqrt{N}$ under the assumption about mean stationarity and that $\sigma_{2\varepsilon}=\sigma_{4\varepsilon}=\sigma_\alpha^2=\tau=1$. For these parameter values, the local power of the test based on J_{II} is always lowest. According to results in Proposition 3 and 5, the local power of the test based on J_I is greater than or equal to that of the test based on J_{II} when $\tau(\tau-1)\leq\sigma_\alpha^2\sigma_{2\varepsilon}/\sigma_{4\varepsilon}$. In Figure 3 the same comparison is done under the assumption about covariance stationarity. As the tests based on t_I, $\bar{t}_I$ and J_I have power against local alternatives where $c = (1-\rho)\sqrt{N}$ and the test based on J_{II} has power against local alternatives where $\tilde{c} = (1-\rho)\,N$, the local power of the tests is shown as a function of $(1-\rho)$ when $T+1$ equals 5 or 10 and N equals 500 or 1000. We see that even for values of ρ very close to unity, the local power of the test based on J_{II} is very high. In particular, the local power of the test based on J_{II} is much higher than that of the other tests. The disadvantage of the test based on J_{II} is that its power depends very much on the assumption being made about the initial values.

Figure 2: Local power of unit root tests based on t_I and $\bar{t}_I$ (solid line), J_I (dashed line) and J_{II} (dotted line) under mean stationarity

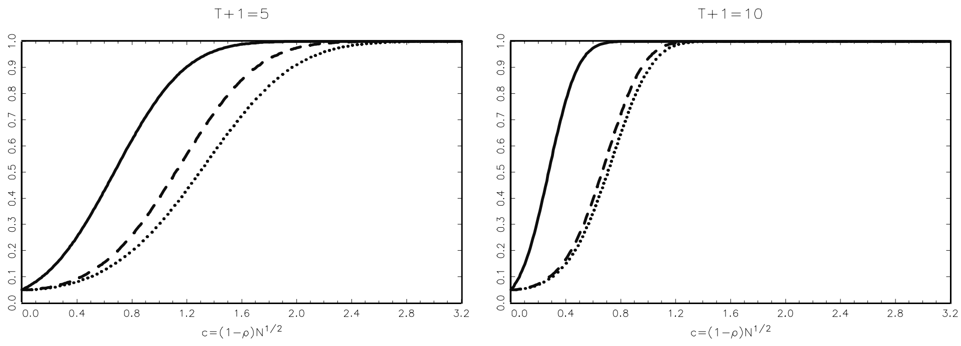
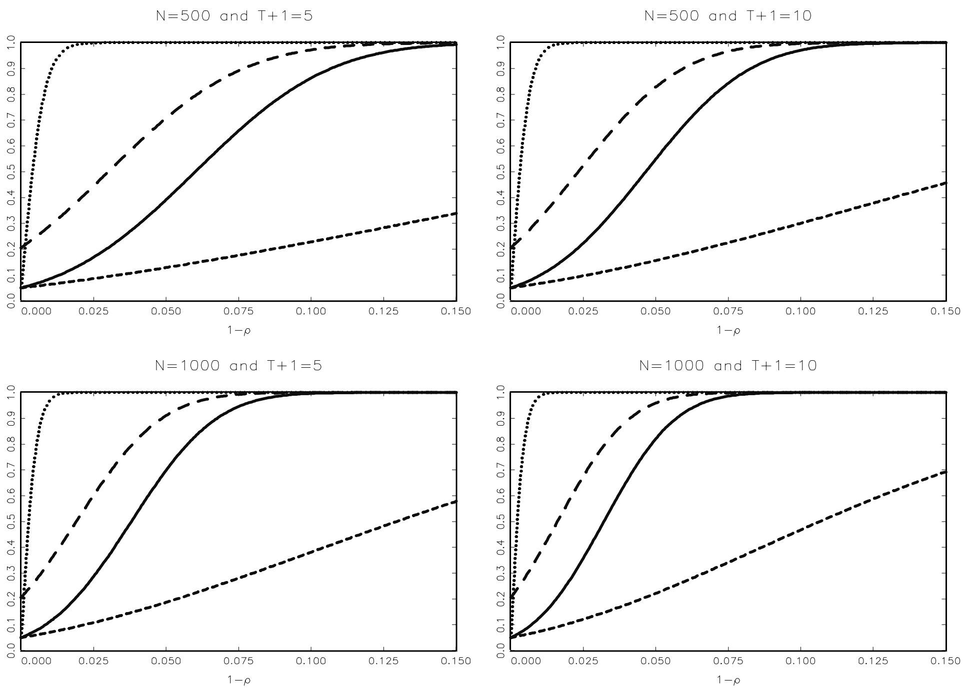


Figure 3: Local power of unit root tests based on t_I (solid line), $\bar{t}_I$ (dashed line), J_I (short-dashed line) and J_{II} (dotted line) under covariance stationarity



4 A simulation study

In this section the analytical results obtained in Section 3 are illustrated in a simulation study. The simulated model is the following

$$y_{i0} = \mathbf{1}_{\{|\rho|<1\}}\alpha_i + \varepsilon_{i0} \tag{42}$$

$$y_{it} = \rho y_{it-1} + (1-\rho)\alpha_i + \varepsilon_{it} \quad \text{for } t = 1, ..., T \tag{43}$$

with

$$\varepsilon_{it} \sim \text{iid}N(0,1) \quad \alpha_i \sim \text{iid}N\left(0, \sigma_\alpha^2\right) \quad \varepsilon_{i0} \sim \text{iid}N(0, \tau(\rho)) \tag{44}$$

The simulations are carried out for different values of T, N and ρ which are $T+1 = 5, 10$, $N = 100, 250, 500, 1000$ and $\rho = 0.900, 0.950, 0.975, 0.990, 1.000$. The results are based on 5000 replications of the model. We consider different simulation setups where either $\tau(\rho) = 1$ when $-1 < \rho \leq 1$ or $\tau(\rho) = 1/\left(1-\rho^2\right)$ when $-1 < \rho < 1$. So we use $\tau(\rho) = 1$ under both the unit root hypothesis and the mean stationary alternative. In addition we consider different simulation setups where the value of σ_α^2 is either 1 or 100. The simulation results with $\sigma_\alpha^2 = 1$ are reported in this section and the simulation results with $\sigma_\alpha^2 = 100$ are reported in Appendix B.

In Table 1 and 2 the results for the statistics based on the Arellano-Bover type moment conditions are summarized. Table 1 corresponds to the unit root case and the mean stationary alternative and Table 2 corresponds to the covariance stationary alternative. The tables show the empirical mean and standard deviation of the Arellano-Bover GMM estimator $\hat{\rho}_I$ and the empirical rejection probabilities of unit root tests based on the t-statistic t_I (one-sided), the normalized coefficient statistic $\bar{t}_I$ (one-sided) and the statistic J_I. The empirical rejection probabilities are all based on tests at the 5% nominal significance level. For comparison the analytical rejection probabilities (i.e. the local power) of the tests are shown in brackets.

In Table 1 we see that the empirical size of all tests is quite close to nominal size of 5%. Although the tests based on t_I and $\bar{t}_I$ have a tendency to reject the null hypothesis too often when $T+1 = 10$ and $N = 100, 250$. The empirical power of the tests is increasing in T and the increase can be quite dramatic when increasing $T+1$ from 5 to 10. In addition, the empirical power of the tests based on t_I and $\bar{t}_I$ is similar and always higher than that of the test based on J_I. These findings are all in accordance with the analytic results in Section 3.1. In Table 2 we see that the empirical power of the test based $\bar{t}_I$ is always higher than that of the test based on t_I. The difference is quite high for most values of T, N and ρ. As in Table 1 the empirical power of the tests based on t_I and $\bar{t}_I$ are always higher than that of the test based on J_I. Comparing the empirical power of the tests in Table 1 and 2, we find that the tests based on t_I and J_I have higher power under the mean stationary alternative than under the covariance stationary alternative. Again these findings are in accordance with the analytical results in Section 3.1. The empirical power of the tests is compared to their local power in Figure 4 (corresponds to Table 1) and Figure 5 (corresponds to Table 2). With a few exceptions, the local power provides a good approximation to the actual power of the tests. The approximation is less accurate for the test based on J_I under mean stationary alternative and the test based on t_I under the covariance stationary alternative when $T + 1 = 10$ and $N = 100, 250$.

In Table 3 and 4 the results for the statistics based on the Arellano-Bond type moment conditions are summarized. Table 3 corresponds to the unit root case and the mean stationary alternative and Table 4 corresponds to the covariance stationary alternative. The tables show the empirical mean and standard deviation of the Arellano-Bond GMM estimator $\hat{\rho}_{II}$ and the empirical rejection probabilities of a unit root test based on the statistic J_{II}. In Table 3, we see that the empirical size of the test is very close to nominal size of 5%. The empirical power of the test is increasing in T but can be quite low for values of ρ close to unity. Comparing with Table 1, we see that under the mean stationary alternative, the empirical power of the test based on J_{II} is much lower than that of the tests based on t_I, $\bar{t}_I$ and J_I. This is in accordance with the analytical results in Section 3.1 and 3.2, compare also with Figure 2. On the other hand, in Table 4 we see that the empirical power of the test based on J_{II} is very high even for values of ρ close to unity. Comparing with Table 2, we see that the test based on J_{II} has much higher power than the tests based on t_I, $\bar{t}_I$ and J_I, compare also with Figure 3. The empirical power of the test is compared to the local power in Figure 6 and again we see that the local power provides a good approximation to the actual power. Turning to the Arellano-Bond GMM estimator $\hat{\rho}_{II}$, in Table 3 the asymptotic bias of the estimator seems to approach -1 as ρ approaches unity. In addition, its variance seems to be constant for all values of ρ and N. On the other hand, in Table 4 the behavior of this estimator seems to be quite different. It still has a downward bias but its variance is decreasing as N increases. These findings have not been investigated in detail in this paper but provides an interesting direction for future research.

Finally, let us shortly summarize the results from the simulation experiment where $\sigma_\alpha^2 = 100$, see Appendix B. The test based on J_I is invariant with respect to this parameter and therefore we do not report its empirical rejection probabilities. The empirical power of all other tests is lower compared to when $\sigma_\alpha^2 = 1$. This is not surprising since all tests rely on the level and the first-differences of the variable y_{it} being correlated. In addition, only the local power of the test based on J_{II} under the assumption about mean stationarity provides a good approximation to its actual power. As explained previously, this is not surprising as the local power of the other tests is only expected to provide a good approximation when the behavior of y_{it} is not dominated by the term α_i. This is obviously not the case when $\sigma_\alpha^2 = 100$. We see that the test based on J_{II} still has very high power under the covariance stationary alternative.

Table 1: Simulation results for statistics based on the Arellano-Bover type moment conditions under mean stationarity with $\tau(\rho) = 1$ and $\sigma_\alpha^2 = 1$

ρ	$T+1$	N	Mean $\hat{\rho}_I$	Std. $\hat{\rho}_I$	$P(t_I < q_1)$ *	$P(\bar{t}_I < q_1)$ *	$P(J_I > q_2)$ **
0.900	5	100	0.9018	0.0524	0.6770 (0.7895)	0.7128 (0.7895)	0.2760 (0.4028)
0.900	5	250	0.9007	0.0320	0.9500 (0.9871)	0.9658 (0.9871)	0.7232 (0.8429)
0.900	5	500	0.9004	0.0223	0.9992 (0.9999)	0.9996 (0.9999)	0.9688 (0.9937)
0.900	5	1000	0.9003	0.0157	1.0000 (1.0000)	1.0000 (1.0000)	1.0000 (1.0000)
0.900	10	100	0.9011	0.0260	0.9994 (1.0000)	0.9992 (1.0000)	0.5548 (0.9349)
0.900	10	250	0.9005	0.0158	1.0000 (1.0000)	1.0000 (1.0000)	0.9988 (1.0000)
0.900	10	500	0.9005	0.0108	1.0000 (1.0000)	1.0000 (1.0000)	1.0000 (1.0000)
0.900	10	1000	0.9004	0.0073	1.0000 (1.0000)	1.0000 (1.0000)	1.0000 (1.0000)
0.950	5	100	0.9517	0.0478	0.3226 (0.3372)	0.3330 (0.3372)	0.0974 (0.1189)
0.950	5	250	0.9507	0.0292	0.5618 (0.6147)	0.5878 (0.6147)	0.2292 (0.2535)
0.950	5	500	0.9504	0.0203	0.8140 (0.8630)	0.8386 (0.8630)	0.4490 (0.4998)
0.950	5	1000	0.9503	0.0144	0.9712 (0.9871)	0.9758 (0.9871)	0.7872 (0.8429)
0.950	10	100	0.9506	0.0222	0.8992 (0.9123)	0.8412 (0.9123)	0.1260 (0.2650)
0.950	10	250	0.9503	0.0135	0.9940 (0.9990)	0.9952 (0.9990)	0.6074 (0.7074)
0.950	10	500	0.9504	0.0092	1.0000 (1.0000)	1.0000 (1.0000)	0.9628 (0.9816)
0.950	10	1000	0.9504	0.0063	1.0000 (1.0000)	1.0000 (1.0000)	1.0000 (1.0000)
0.975	5	100	0.9766	0.0457	0.1666 (0.1509)	0.1662 (0.1509)	0.0600 (0.0651)
0.975	5	250	0.9757	0.0279	0.2408 (0.2493)	0.2544 (0.2493)	0.0888 (0.0905)
0.975	5	500	0.9754	0.0194	0.3626 (0.3914)	0.3858 (0.3914)	0.1364 (0.1392)
0.975	5	1000	0.9752	0.0137	0.5722 (0.6147)	0.5936 (0.6147)	0.2504 (0.2535)
0.975	10	100	0.9754	0.0204	0.5494 (0.4424)	0.4284 (0.4424)	0.0588 (0.0868)
0.975	10	250	0.9752	0.0125	0.7488 (0.7663)	0.7250 (0.7663)	0.1610 (0.1645)
0.975	10	500	0.9753	0.0084	0.9306 (0.9563)	0.9342 (0.9563)	0.3402 (0.3405)
0.975	10	1000	0.9754	0.0058	0.9978 (0.9990)	0.9978 (0.9990)	0.6906 (0.7074)
0.990	5	100	0.9916	0.0445	0.0988 (0.0808)	0.0944 (0.0808)	0.0500 (0.0523)
0.990	5	250	0.9907	0.0272	0.1114 (0.1043)	0.1126 (0.1043)	0.0582 (0.0559)
0.990	5	500	0.9904	0.0189	0.1356 (0.1363)	0.1406 (0.1363)	0.0616 (0.0620)
0.990	5	1000	0.9903	0.0134	0.1946 (0.1921)	0.2024 (0.1921)	0.0742 (0.0749)
0.990	10	100	0.9903	0.0194	0.2732 (0.1480)	0.1722 (0.1480)	0.0478 (0.0551)
0.990	10	250	0.9902	0.0119	0.2932 (0.2431)	0.2558 (0.2432)	0.0718 (0.0633)
0.990	10	500	0.9903	0.0080	0.3798 (0.3809)	0.3616 (0.3809)	0.0842 (0.0785)
0.990	10	1000	0.9903	0.0055	0.5668 (0.5997)	0.5688 (0.5997)	0.1254 (0.1147)
1.000	5	100	1.0015	0.0438	0.0674 (0.0500)	0.0606 (0.0500)	0.0478 (0.0500)
1.000	5	250	1.0006	0.0267	0.0544 (0.0500)	0.0520 (0.0500)	0.0532 (0.0500)
1.000	5	500	1.0004	0.0186	0.0528 (0.0500)	0.0546 (0.0500)	0.0522 (0.0500)
1.000	5	1000	1.0002	0.0132	0.0500 (0.0500)	0.0510 (0.0500)	0.0508 (0.0500)
1.000	10	100	1.0002	0.0188	0.1318 (0.0500)	0.0740 (0.0500)	0.0444 (0.0500)
1.000	10	250	1.0002	0.0115	0.0866 (0.0500)	0.0702 (0.0500)	0.0592 (0.0500)
1.000	10	500	1.0003	0.0078	0.0624 (0.0500)	0.0574 (0.0500)	0.0572 (0.0500)
1.000	10	1000	1.0003	0.0053	0.0520 (0.0500)	0.0494 (0.0500)	0.0626 (0.0500)

* q_1 is the 5%-quantile of the standard normal distribution
** q_2 is the 95%-quantile of the χ^2-distribution with $\frac{1}{2}T(T-1)$ degrees of freedom
The numbers in column 6-8 are the empirical rejection probabilities and the local power (in brackets)

Table 2: Simulation results for statistics based on the Arellano-Bover type moment conditions under covariance stationarity with $\tau(\rho) = 1/(1-\rho^2)$ and $\sigma_\alpha^2 = 1$

ρ	$T+1$	N	Mean $\hat{\rho}_I$	Std. $\hat{\rho}_I$	$P(t_I < q_1)$ *	$P(\bar{t}_I < q_1)$ *	$P(J_I > q_2)$ **
0.900	5	100	0.8971	0.0975	0.3812 (0.3372)	0.6328 (0.6563)	0.1066 (0.1189)
0.900	5	250	0.8993	0.0560	0.6310 (0.6147)	0.8558 (0.8674)	0.2618 (0.2535)
0.900	5	500	0.8999	0.0386	0.8634 (0.8630)	0.9700 (0.9723)	0.5010 (0.4998)
0.900	5	1000	0.9002	0.0270	0.9836 (0.9871)	0.9988 (0.9989)	0.8384 (0.8429)
0.900	10	100	0.9006	0.0399	0.9558 (0.9123)	0.9780 (0.9853)	0.1460 (0.2650)
0.900	10	250	0.9005	0.0246	0.9986 (0.9990)	0.9998 (1.0000)	0.6900 (0.7074)
0.900	10	500	0.9007	0.0170	1.0000 (1.0000)	1.0000 (1.0000)	0.9812 (0.9816)
0.900	10	1000	0.9005	0.0117	1.0000 (1.0000)	1.0000 (1.0000)	1.0000 (1.0000)
0.950	5	100	0.9425	0.1062	0.1716 (0.1509)	0.4028 (0.4168)	0.0618 (0.0651)
0.950	5	250	0.9483	0.0577	0.2612 (0.2493)	0.5546 (0.5580)	0.0946 (0.0905)
0.950	5	500	0.9497	0.0383	0.3914 (0.3914)	0.7018 (0.7078)	0.1420 (0.1392)
0.950	5	1000	0.9501	0.0268	0.6110 (0.6147)	0.8642 (0.8674)	0.2576 (0.2535)
0.950	10	100	0.9490	0.0365	0.6672 (0.4424)	0.7442 (0.7510)	0.0586 (0.0868)
0.950	10	250	0.9498	0.0231	0.8202 (0.7663)	0.9336 (0.9393)	0.1708 (0.1645)
0.950	10	500	0.9503	0.0161	0.9576 (0.9563)	0.9920 (0.9943)	0.3668 (0.3405)
0.950	10	1000	0.9504	0.0111	0.9984 (0.9990)	1.0000 (1.0000)	0.7312 (0.7074)
0.975	5	100	0.9631	0.1115	0.0986 (0.0903)	0.2834 (0.3028)	0.0532 (0.0536)
0.975	5	250	0.9716	0.0618	0.1200 (0.1229)	0.3626 (0.3676)	0.0614 (0.0593)
0.975	5	500	0.9744	0.0389	0.1682 (0.1685)	0.4384 (0.4452)	0.0700 (0.0692)
0.975	5	1000	0.9750	0.0270	0.2542 (0.2493)	0.5492 (0.5580)	0.0922 (0.0905)
0.975	10	100	0.9733	0.0328	0.4098 (0.1854)	0.4628 (0.4711)	0.0458 (0.0581)
0.975	10	250	0.9744	0.0220	0.4370 (0.3231)	0.6378 (0.6419)	0.0788 (0.0716)
0.975	10	500	0.9751	0.0155	0.5710 (0.5128)	0.7968 (0.8036)	0.1000 (0.0979)
0.975	10	1000	0.9754	0.0107	0.7770 (0.7663)	0.9364 (0.9393)	0.1766 (0.1645)
0.990	5	100	0.9770	0.1037	0.0560 (0.0640)	0.1822 (0.2420)	0.0496 (0.0506)
0.990	5	250	0.9833	0.0775	0.0518 (0.0734)	0.2418 (0.2647)	0.0546 (0.0514)
0.990	5	500	0.9873	0.0515	0.0638 (0.0852)	0.2798 (0.2917)	0.0514 (0.0529)
0.990	5	1000	0.9899	0.0281	0.0896 (0.1043)	0.3268 (0.3317)	0.0560 (0.0559)
0.990	10	100	0.9884	0.0267	0.2582 (0.0893)	0.2262 (0.3007)	0.0430 (0.0512)
0.990	10	250	0.9892	0.0197	0.2000 (0.1209)	0.3356 (0.3639)	0.0626 (0.0531)
0.990	10	500	0.9899	0.0147	0.2070 (0.1650)	0.4258 (0.4397)	0.0602 (0.0564)
0.990	10	1000	0.9903	0.0103	0.2624 (0.2432)	0.5270 (0.5502)	0.0758 (0.0633)

* q_1 is the 5%-quantile of the standard normal distribution
** q_2 is the 95%-quantile of the χ^2-distribution with $\frac{1}{2}T(T-1)$ degrees of freedom
The numbers in column 6-8 are the empirical rejection probabilities and the local power (in brackets)

Figure 4: Comparison of actual and local power of the tests based on the Arellano-Bover type moment conditions under mean stationarity with $\tau(\rho) = 1$ and $\sigma_\alpha^2 = 1$. (Corresponds to Table 1)

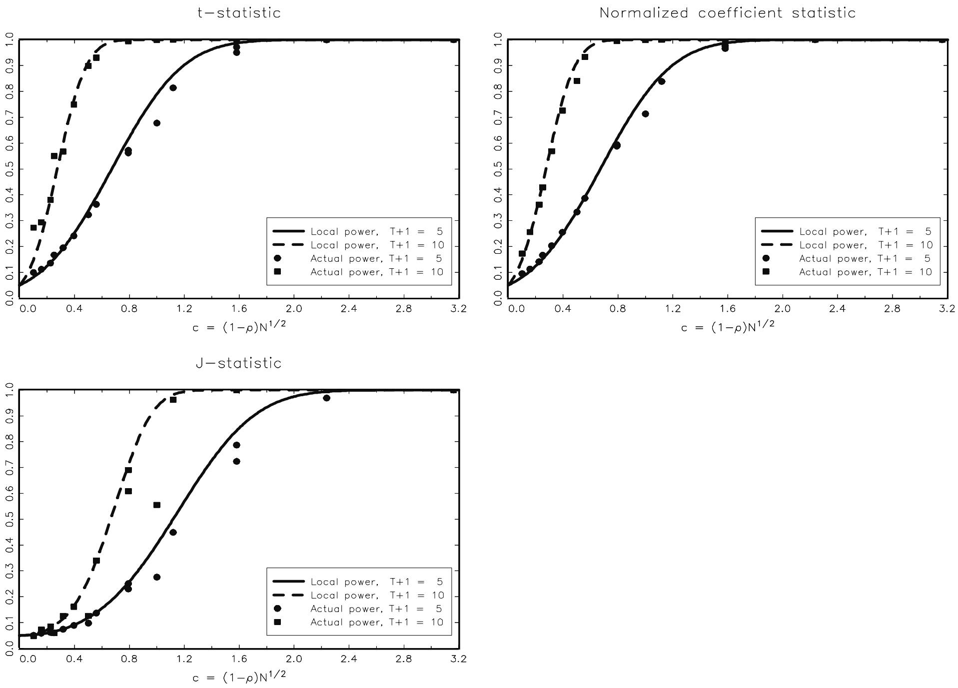


Figure 5: Comparison of actual and local power of the tests based on the Arellano-Bover type moment conditions under covariance stationarity with $\tau(\rho) = 1/(1-\rho^2)$ and $\sigma_\alpha^2 = 1$. (Corresponds to Table 2)

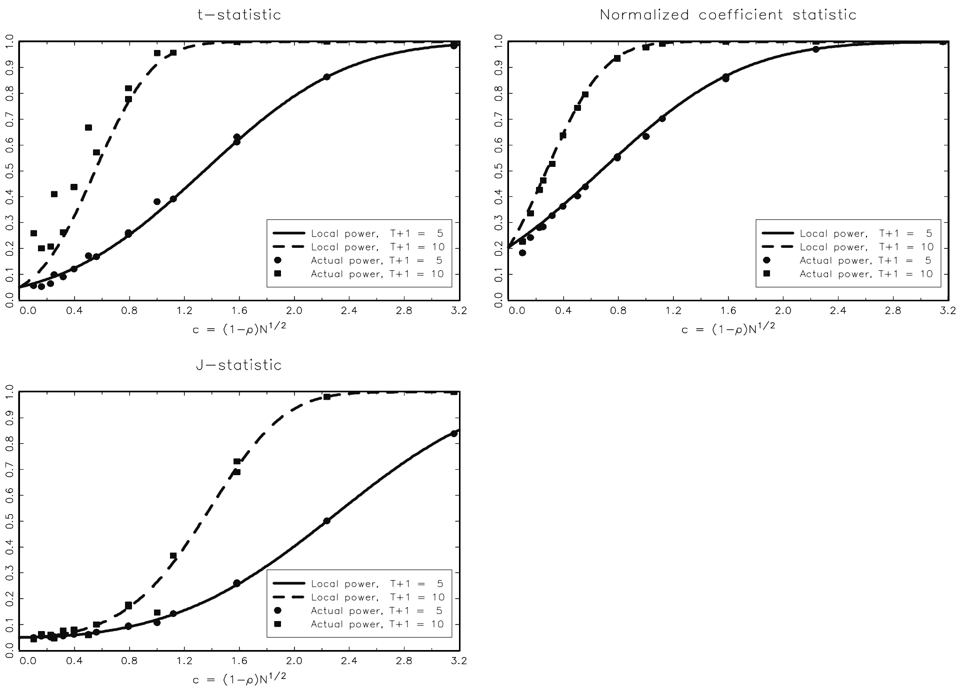


Table 3: Simulation results for statistics based on the Arellano-Bond type moment conditions under mean stationarity with $\tau(\rho) = 1$ and $\sigma_\alpha^2 = 1$

ρ	$T+1$	N	Mean $(\hat{\rho}_{II} - \rho)$	Std. $(\hat{\rho}_{II} - \rho)$	$P(J_{II} > q_1)$*
0.900	5	100	-0.6017	0.3960	0.2526 (0.3028)
0.900	5	250	-0.3722	0.3103	0.6568 (0.7050)
0.900	5	500	-0.2177	0.2336	0.9414 (0.9648)
0.900	5	1000	-0.1202	0.1664	0.9996 (0.9999)
0.900	10	100	-0.6412	0.1474	0.5602 (0.8928)
0.900	10	250	-0.4174	0.1135	0.9986 (1.0000)
0.900	10	500	-0.2652	0.0843	1.0000 (1.0000)
0.900	10	1000	-0.1531	0.0589	1.0000 (1.0000)
0.950	5	100	-0.8570	0.4389	0.0902 (0.0997)
0.950	5	250	-0.7063	0.4201	0.1970 (0.1942)
0.950	5	500	-0.5324	0.3752	0.3638 (0.3779)
0.950	5	1000	-0.3511	0.2989	0.6922 (0.7050)
0.950	10	100	-0.8529	0.1674	0.1246 (0.2333)
0.950	10	250	-0.6993	0.1546	0.5732 (0.6361)
0.950	10	500	-0.5391	0.1350	0.9492 (0.9620)
0.950	10	1000	-0.3686	0.1046	0.9998 (1.0000)
0.975	5	100	-0.9605	0.4354	0.0578 (0.0612)
0.975	5	250	-0.9046	0.4357	0.0840 (0.0796)
0.975	5	500	-0.8250	0.4301	0.1156 (0.1140)
0.975	5	1000	-0.6952	0.4181	0.1942 (0.1942)
0.975	10	100	-0.9540	0.1681	0.0566 (0.0821)
0.975	10	250	-0.8936	0.1664	0.1556 (0.1482)
0.975	10	500	-0.8087	0.1656	0.3106 (0.2980)
0.975	10	1000	-0.6782	0.1536	0.6456 (0.6361)
0.990	5	100	-0.9958	0.4354	0.0514 (0.0517)
0.990	5	250	-0.9838	0.4344	0.0576 (0.0544)
0.990	5	500	-0.9665	0.4287	0.0622 (0.0589)
0.990	5	1000	-0.9363	0.4361	0.0682 (0.0683)
0.990	10	100	-0.9919	0.1656	0.0446 (0.0545)
0.990	10	250	-0.9815	0.1642	0.0740 (0.0617)
0.990	10	500	-0.9629	0.1664	0.0900 (0.0749)
0.990	10	1000	-0.9266	0.1664	0.1206 (0.1060)
1.000	5	100	-1.0042	0.4397	0.0500 (0.0500)
1.000	5	250	-1.0030	0.4332	0.0566 (0.0500)
1.000	5	500	-1.0007	0.4280	0.0532 (0.0500)
1.000	5	1000	-1.0040	0.4276	0.0516 (0.0500)
1.000	10	100	-1.0004	0.1647	0.0412 (0.0500)
1.000	10	250	-1.0026	0.1635	0.0586 (0.0500)
1.000	10	500	-1.0028	0.1644	0.0610 (0.0500)
1.000	10	1000	-1.0009	0.1631	0.0618 (0.0500)

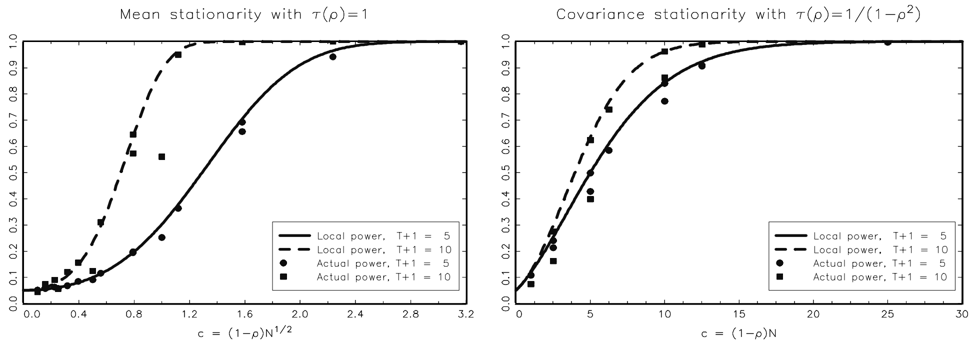
* q_1 is the 95%-quantile of the χ^2-distribution with $\frac{1}{2}T(T-1)$ degrees of freedom
The numbers in column 6 are the empirical rejection probabilities and the local power (in brackets)

Table 4: Simulation results for statistics based on the Arellano-Bond type moment conditions under covariance stationarity with $\tau(\rho) = 1/\left(1-\rho^2\right)$ and $\sigma_\alpha^2 = 1$

ρ	$T+1$	N	Mean $(\hat{\rho}_{II} - \rho)$	Std. $(\hat{\rho}_{II} - \rho)$	$P(J_{II} > q_1)$*
0.900	5	100	-0.2966	0.2582	0.7720 (0.8429)
0.900	5	250	-0.1440	0.1640	0.9976 (0.9991)
0.900	5	500	-0.0755	0.1155	1.0000 (1.0000)
0.900	5	1000	-0.0392	0.0794	1.0000 (1.0000)
0.900	10	100	-0.5003	0.1266	0.8614 (0.9620)
0.900	10	250	-0.2864	0.0862	1.0000 (1.0000)
0.900	10	500	-0.1685	0.0593	1.0000 (1.0000)
0.900	10	1000	-0.0921	0.0400	1.0000 (1.0000)
0.950	5	100	-0.4588	0.3368	0.4280 (0.4998)
0.950	5	250	-0.2485	0.2270	0.9062 (0.9230)
0.950	5	500	-0.1371	0.1574	0.9992 (0.9991)
0.950	5	1000	-0.0725	0.1075	1.0000 (1.0000)
0.950	10	100	-0.6599	0.1519	0.3984 (0.6361)
0.950	10	250	-0.4361	0.1154	0.9886 (0.9915)
0.950	10	500	-0.2799	0.0832	1.0000 (1.0000)
0.950	10	1000	-0.1626	0.0557	1.0000 (1.0000)
0.975	5	100	-0.6353	0.4052	0.2134 (0.2535)
0.975	5	250	-0.4021	0.3100	0.5844 (0.6104)
0.975	5	500	-0.2420	0.2225	0.9084 (0.9230)
0.975	5	1000	-0.1337	0.1517	0.9990 (0.9991)
0.975	10	100	-0.7928	0.1666	0.1614 (0.2980)
0.975	10	250	-0.6031	0.1448	0.7396 (0.7694)
0.975	10	500	-0.4324	0.1159	0.9908 (0.9915)
0.975	10	1000	-0.2748	0.0816	1.0000 (1.0000)
0.990	5	100	-0.8166	0.4343	0.1074 (0.1189)
0.990	5	250	-0.6379	0.4010	0.2400 (0.2535)
0.990	5	500	-0.4562	0.3373	0.4980 (0.4998)
0.990	5	1000	-0.2844	0.2461	0.8404 (0.8429)
0.990	10	100	-0.9052	0.1704	0.0736 (0.1238)
0.990	10	250	-0.7910	0.1664	0.2748 (0.2980)
0.990	10	500	-0.6548	0.1546	0.6236 (0.6361)
0.990	10	1000	-0.4844	0.1268	0.9626 (0.9620)

* q_1 is the 95%-quantile of the χ^2-distribution with $\frac{1}{2}T(T-1)$ degrees of freedom
The numbers in column 6 are the empirical rejection probabilities and the local power (in brackets)

Figure 6: Comparison of actual and local power of the test based on the statistic J_{II} with $\sigma_\alpha^2 = 1$. (Corresponds to Table 3 and 4)



5 Conclusions

In this paper we have considered GMM-based unit root inference in an autoregressive panel data model with individual-specific levels. More specifically, we have investigated the performance of various GMM-based unit root tests in terms of their local power. We find that a unit root test based on the Arellano-Bover GMM estimator performs well and is asymptotically equivalent to the unit root test suggested by Breitung & Meyer (1994). On the other hand, the Arellano-Bond GMM estimator can not be used for unit root inference. Instead a moment condition test of the hypothesis that the Arellano-Bond type moment conditions do not identify the AR parameter is valid as a unit root test. Under the covariance stationary alternative, the local power of this test is strikingly high even for values of the AR parameter very close to unity. In particular, this moment condition test clearly outperforms the test based on the Arellano-Bover GMM estimator. Under the mean stationary alternative, the situation is completely different. In this case, the local power of the moment condition test is likely to be quite low. This result illustrates that the underlying assumptions concerning the initial values are important.

The results concerning the Arellano-Bond GMM estimator are not only of interest in relation to unit root inference. It is well-known that the usual asymptotic representation of the Arellano-Bond GMM estimator does not provide a good approximation to its actual behavior for high values of the AR parameter. Simulation studies show that the estimator has a downward bias and a large variance even in large samples. Therefore, it would be interesting to see if the type of asymptotic representations derived in this paper can explain these findings.

Finally, Blundell & Bond (1998) and Blundell, Bond & Windmeijer (2000) suggest using a GMM estimator based on all of the Arellano-Bond type moment conditions and the additional Arellano-Bover type moment conditions which are not redundant. When the AR parameter is high, the results in this paper suggest that the GMM estimator based on all of the Arellano-Bover type moment conditions and the additional Arellano-Bond type moment conditions is more efficient. Hence, it should be used instead.

A Appendices

This appendix contains the proofs of the propositions in Section 3. They are all based on standard asymptotic theory which in this case is Markov's Law of Large Numbers and the Liapounov Central Limit Theorem, see for example White (2001).

In the following I_k denotes an identity matrix of dimension k and ι_k denotes a $k \times 1$ vector of ones. In addition we use that $m = \frac{1}{2}T(T-1)$.

Throughout this appendix the expressions for y_{it} and Δy_{it} given below are used. By recursive substitution in (1) and by insertion of the expression for the initial value y_{i0} given in Assumption 2, the following expression for y_{it} when $-1 < \rho \leq 1$ is obtained

$$y_{it} = \mathbf{1}_{\{|\rho|<1\}}\alpha_i + \rho^t \sqrt{\tau(\rho)}\varepsilon_{i0} + \rho^{t-1}\varepsilon_{i1} + ... + \varepsilon_{it} \quad \text{for } t = 0, ..., T \tag{45}$$

Using this yields the following expression for Δy_{it}

$$\Delta y_{it} = (\rho - 1)\left(\rho^{t-1}\sqrt{\tau(\rho)}\varepsilon_{i0} + \rho^{t-2}\varepsilon_{i1} + ... + \varepsilon_{it-1}\right) + \varepsilon_{it} \quad \text{for } t = 1, ..., T \tag{46}$$

A.1 Preliminary results

Lemma 1 *Under the local-to-unity sequence for ρ given by $\rho = 1 - c/N^k$ for $c, k > 0$ the following hold*

$$\rho^t = 1 - t\frac{c}{N^k} + o\left(N^{-k}\right) \tag{47}$$

$$\frac{1}{1-\rho^2} = \frac{N^k}{2c} + o\left(N^{-k}\right) \tag{48}$$

Proof: The binomial formula yields

$$\rho^t = \left(1 - \frac{c}{N^k}\right)^t = 1 - t\frac{c}{N^k} + \frac{t(t-1)}{2!}\frac{c^2}{N^{2k}} - \frac{t(t-1)(t-2)}{3!}\frac{c^3}{N^{3k}} + ... + \frac{(-c)^t}{N^{tk}}$$

and the results follow directly. □

Lemma 2 *Under Assumption 1, 2 and 3 the following results hold*

$$\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{it}^2 \xrightarrow{P} \sigma_{2\varepsilon} \quad \textit{as } N \to \infty \quad \textit{for all } t = 0, ..., T \tag{49}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\varepsilon_{it}\varepsilon_{is} \xrightarrow{w} N(0, \sigma_{4\varepsilon}) \quad \textit{as } N \to \infty \quad \textit{for all } t, s = 0, ..., T \textit{ with } s \neq t \tag{50}$$

Proof:

Markov's Law of Large Numbers can be applied to the sequence in (49) since $E|\varepsilon_{it}|^{4+\delta} < K$ for some $\delta > 0$ and all $i = 1, ..., N$, $t = 0, ..., T$. Thus, we have $\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{it}^2 - \frac{1}{N}\sum_{i=1}^{N}E\left(\varepsilon_{it}^2\right) \xrightarrow{P} 0$ as $N \to \infty$ which in combination with $\frac{1}{N}\sum_{i=1}^{N}E\left(\varepsilon_{it}^2\right) = \frac{1}{N}\sum_{i=1}^{N}\sigma_{i\varepsilon}^2 \to \sigma_{2\varepsilon}$ as $N \to \infty$ yields the result in (49). By using similar arguments the Liapounov Central Limit Theorem can be applied to the sequence in (50). Also we have that ε_{it} and ε_{is} for $t = 0, ..., T$ and $t \neq s$ are independent such that $E(\varepsilon_{it}\varepsilon_{is}) = 0$ and $E\left(\varepsilon_{it}^2\varepsilon_{is}^2\right) = \sigma_{i\varepsilon}^4$ which altogether give the result in (50). □

A.2 Proofs of the propositions in Section 3.1

For repetition the regression model is the following

$$\begin{aligned} y_{it} &= \rho y_{it-1} + v_{it} \\ v_{it} &= (1-\rho)\,\alpha_i + \varepsilon_{it} \end{aligned} \qquad \text{for } i = 1, ..., N \text{ and } t = 2, ..., T \tag{51}$$

Using stacked notation the regressor is $y_{i,-1} = (y_{i1}, ..., y_{iT-1})'$, the regression error is $v_i = (v_{i2}, ..., v_{iT})'$ and the $(T-1) \times m$ matrix of instruments is

$$Z_{i1} = \begin{bmatrix} \Delta y_{i1} & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \Delta y_{i1} & \Delta y_{i2} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & \Delta y_{i1} & \cdots & \Delta y_{iT-1} \end{bmatrix} \tag{52}$$

The standard IV estimator $\tilde{\rho}_I$ corresponding to this instrumental variables regression can be expressed in the following way

$$\tilde{\rho}_I = \rho + \left(\sum_{i=1}^{N} y'_{i,-1} Z_{i1} \left(\sum_{i=1}^{N} Z'_{i1} Z_{i1} \right)^{-1} \sum_{i=1}^{N} Z'_{i1} y_{i,-1} \right)^{-1} \sum_{i=1}^{N} y'_{i,-1} Z_{i1} \left(\sum_{i=1}^{N} Z'_{i1} Z_{i1} \right)^{-1} \sum_{i=1}^{N} Z'_{i1} v_i \tag{53}$$

The asymptotic properties of this estimator is given by the results in Lemma 3 below.

Lemma 3 *Under Assumption 1, 2, 3 and the local-to-unity sequence for ρ given by $\rho = 1 - c/\sqrt{N}$ for $c \geq 0$, the following results hold*

$$(i) \text{ and } c \geq 0: \quad \frac{1}{N} \sum_{i=1}^{N} Z'_{i1} y_{i,-1} \xrightarrow{P} \sigma_{2\varepsilon} \iota_m \quad \text{as } N \to \infty \tag{a1}$$

$$(ii) \text{ and } c > 0: \quad \frac{1}{N} \sum_{i=1}^{N} Z'_{i1} y_{i,-1} \xrightarrow{P} \frac{1}{2} \sigma_{2\varepsilon} \iota_m \quad \text{as } N \to \infty \tag{a2}$$

$$\frac{1}{N} \sum_{i=1}^{N} Z'_{i1} Z_{i1} \xrightarrow{P} \sigma_{2\varepsilon} I_m \quad \text{as } N \to \infty \tag{b}$$

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} Z'_{i1} v_i \xrightarrow{w} N\left(0, \sigma_{4\varepsilon} I_m\right) \quad \text{as } N \to \infty \tag{c}$$

In particular, the limiting distribution of the IV estimator $\tilde{\rho}_I$ is given by

$$(i) \text{ and } c \geq 0: \quad \sqrt{N}\left(\tilde{\rho}_I - \rho\right) \xrightarrow{w} N\left(0, \frac{\sigma_{4\varepsilon}}{\sigma_{2\varepsilon}^2} \frac{2}{T(T-1)}\right) \quad \text{as } N \to \infty \tag{d1}$$

$$(ii) \text{ and } c > 0: \quad \sqrt{N}\left(\tilde{\rho}_I - \rho\right) \xrightarrow{w} N\left(0, 4\frac{\sigma_{4\varepsilon}}{\sigma_{2\varepsilon}^2} \frac{2}{T(T-1)}\right) \quad \text{as } N \to \infty \tag{d2}$$

Proof of Lemma 3:

(a1) and (a2) We prove the results by showing that for $t = 1, ..., T-1$ and $s = 1, ..., t$ the following hold

$$(i) \text{ and } c \geq 0: \quad \frac{1}{N} \sum_{i=1}^{N} \Delta y_{is} y_{it} \xrightarrow{P} \sigma_{2\varepsilon} \text{ as } N \to \infty \tag{54}$$

$$(ii) \text{ and } c > 0: \quad \frac{1}{N} \sum_{i=1}^{N} \Delta y_{is} y_{it} \xrightarrow{P} \frac{1}{2} \sigma_{2\varepsilon} \text{ as } N \to \infty \tag{55}$$

This proves the results in (a1) and (a2) as $\frac{1}{N}\sum_{i=1}^{N} Z_{i1}' y_{i,-1}$ has elements of the form $\frac{1}{N}\sum_{i=1}^{N} \Delta y_{is} y_{it}$ where $t = 1, ..., T-1$ and $s = 1, ..., t$. To show the results in (54) and (55) we use the following expression

$$\Delta y_{is} y_{it} = Q_{1i} Q_{2i} + (\Delta y_{is} - Q_{1i}) y_{it} + Q_{1i} (y_{it} - Q_{2i}) \tag{56}$$

where

$$\begin{aligned} Q_{1i} &= \varepsilon_{is} + (\rho - 1) \rho^{s-1} \sqrt{\tau(\rho)} \varepsilon_{i0} && (57) \\ Q_{2i} &= \rho^{t-s} \varepsilon_{is} + \rho^{t} \sqrt{\tau(\rho)} \varepsilon_{i0} && (58) \end{aligned}$$

and we show that

$$\begin{aligned} (i) \text{ and } c \geq 0: \quad & \frac{1}{N}\sum_{i=1}^{N} Q_{1i} Q_{2i} \xrightarrow{P} \sigma_{2\varepsilon} \quad \text{as } N \to \infty && (59) \\ (ii) \text{ and } c > 0: \quad & \frac{1}{N}\sum_{i=1}^{N} Q_{1i} Q_{2i} \xrightarrow{P} \frac{1}{2}\sigma_{2\varepsilon} \quad \text{as } N \to \infty && (60) \\ & \frac{1}{N}\sum_{i=1}^{N} (\Delta y_{is} - Q_{1i}) y_{it} \xrightarrow{P} 0 \quad \text{as } N \to \infty && (61) \\ & \frac{1}{N}\sum_{i=1}^{N} Q_{1i} (y_{it} - Q_{2i}) \xrightarrow{P} 0 \quad \text{as } N \to \infty && (62) \end{aligned}$$

To show (59) we use the results below which follow by Lemma 1.

$$\begin{aligned} \rho^{t} &\to 1 \quad \text{as } N \to \infty && (63) \\ (\rho - 1) \rho^{t} &\to 0 \quad \text{as } N \to \infty && (64) \\ \frac{\tau(\rho)^{\frac{1}{2}}}{\sqrt{N}} &\to 0 \quad \text{as } N \to \infty && (65) \\ (\rho - 1) \tau(\rho) &\to \begin{cases} 0 & \text{under } (i) \text{ and } c \geq 0 \\ -\frac{1}{2} & \text{under } (ii) \text{ and } c > 0 \end{cases} \quad \text{as } N \to \infty && (66) \end{aligned}$$

We have that

$$\begin{aligned} & \frac{1}{N}\sum_{i=1}^{N} Q_{1i} Q_{2i} \\ = \; & \rho^{t-s} \frac{1}{N}\sum_{i=1}^{N} \varepsilon_{is}^{2} + (\rho - 1) \tau(\rho) \rho^{t+s-1} \frac{1}{N}\sum_{i=1}^{N} \varepsilon_{i0}^{2} + \left(\rho^{t} + (\rho - 1) \rho^{t-1}\right) \frac{\tau(\rho)^{\frac{1}{2}}}{\sqrt{N}} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \varepsilon_{i0} \varepsilon_{is} \end{aligned}$$

According to Lemma 2 and the results above we have that as $N \to \infty$, the first term on the right hand side in the expression above converges in probability to $\sigma_{2\varepsilon}$, the second term converges in probability to zero under (i) and to $-\frac{1}{2}\sigma_{2\varepsilon}$ under (ii), and the third term converges in probability to zero. This proves the result in (59).

The results below are used in the following.

$$E\left(y_{it}^2\right) \leq \sigma_\alpha^2 + \left(\rho^{2t}\tau\left(\rho\right) + \rho^{2(t-1)} + ... + \rho^2 + 1\right)\sigma_{i\varepsilon}^2 = O\left(N^{\frac{1}{2}}\right) \tag{67}$$

$$E\left(\left(\Delta y_{is} - Q_{1i}\right)^2\right) = \left(\left(\rho-1\right)^2\left(\rho^{2(s-2)} + ... + \rho^2 + 1\right)\right)\sigma_{i\varepsilon}^2 = O\left(N^{-1}\right) \tag{68}$$

$$E\left(\left(y_{it} - Q_{2i}\right)^2\right) \leq \sigma_\alpha^2 + \left(\rho^{2(t-1)} + ... + \rho^2 + 1\right)\sigma_{i\varepsilon}^2 = O\left(1\right) \tag{69}$$

$$E\left(Q_{1i}^2\right) = \left(1 + \left(\rho-1\right)^2\rho^{2(s-1)}\tau\left(\rho\right)\right)\sigma_{i\varepsilon}^2 = O\left(1\right) \tag{70}$$

$$E\left(\Delta y_{is}^2\right) = \left(1 + \left(\rho-1\right)^2\left(\rho^{2(s-1)}\tau\left(\rho\right) + \rho^{2(s-2)} + ... + 1\right)\right)\sigma_{i\varepsilon}^2 = O\left(1\right) \tag{71}$$

$$E\left(\left(\Delta y_{is} - \varepsilon_{is}\right)^2\right) = \left(\rho-1\right)^2\left(\rho^{2(s-1)}\tau\left(\rho\right) + \rho^{2(s-2)} + ... + 1\right)\sigma_{i\varepsilon}^2 = O\left(N^{-\frac{1}{2}}\right) \tag{72}$$

$$E\left(v_{it}^2\right) = \left(1-\rho\right)^2\sigma_\alpha^2 + \sigma_{i\varepsilon}^2 = O\left(1\right) \tag{73}$$

They are obtained by using the expressions for y_{it} and Δy_{it} given in (45) and (46) and that all terms in these expressions are independent of each other. In addition the results in Lemma 1 are applied. The result in (61) holds as

$$E\left|\frac{1}{N}\sum_{i=1}^{N}\left(\Delta y_{is} - Q_{1i}\right)y_{it}\right| \leq \frac{1}{N}\sum_{i=1}^{N}\sqrt{E\left(\left(\Delta y_{is} - Q_{1i}\right)^2\right)E\left(y_{it}^2\right)} \leq O\left(N^{-\frac{1}{4}}\right) \tag{74}$$

The first inequality results from the triangle inequality and the Cauchy-Schwarz inequality. The second inequality holds by using that $E\left(y_{it}^2\right) \leq O\left(N^{\frac{1}{2}}\right)$ and $E\left(\left(\Delta y_{is} - Q_{1i}\right)^2\right) = O\left(N^{-1}\right)$ according to the results in (67) and (68). This shows that $E\left|\frac{1}{N}\sum_{i=1}^{N}\left(\Delta y_{is} - Q_{1i}\right)y_{it}\right| \to 0$ as $N \to \infty$ such that $\frac{1}{N}\sum_{i=1}^{N}\left(\Delta y_{is} - Q_{1i}\right)y_{it} \xrightarrow{P} 0$ as $N \to \infty$. The result in (62) holds since

$$E\left|\frac{1}{N}\sum_{i=1}^{N}Q_{1i}\left(y_{it} - Q_{2i}\right)\right|^2 = \frac{1}{N^2}\sum_{i=1}^{N}E\left(Q_{1i}^2\right)E\left(\left(y_{it} - Q_{2i}\right)^2\right) \leq \frac{1}{N}O\left(1\right) \tag{75}$$

The first equality sign results from Q_{1i} and $\left(y_{it} - Q_{2i}\right)$ being independent with means zero such that $Q_{1i}\left(y_{it} - Q_{2i}\right)$ is independent across i with mean zero. The second inequality follows by using that $E\left(\left(y_{it} - Q_{2i}\right)^2\right) \leq O\left(1\right)$ and $E\left(Q_{1i}^2\right) = O\left(1\right)$ according to the results in (69) and (70). This shows that $E\left|\frac{1}{N}\sum_{i=1}^{N}Q_{1i}\left(y_{it} - Q_{2i}\right)\right|^2 \to 0$ as $N \to \infty$ such that $\frac{1}{N}\sum_{i=1}^{N}Q_{1i}\left(y_{it} - Q_{2i}\right) \xrightarrow{P} 0$ as $N \to \infty$. Altogether, we have obtained the desired limits and the results in (a1) and (a2) are proved.

(b) $\frac{1}{N}\sum_{i=1}^{N}Z_{i1}'Z_{i1}$ has non-zero elements of the form $\frac{1}{N}\sum_{i=1}^{N}\Delta y_{is}\Delta y_{it}$ for $t = 1, ..., T-1$ and $s = 1, ..., t$ where the diagonal elements correspond to $t = s$. We prove the result by showing that

$$\frac{1}{N}\sum_{i=1}^{N}\Delta y_{is}\Delta y_{it} \xrightarrow{P} \begin{cases} \sigma_{2\varepsilon} & \text{for } t = s \\ 0 & \text{for } t \neq s \end{cases} \quad \text{as } N \to \infty \tag{76}$$

This is done by showing that for all $t, s = 1, ..., T-1$

$$\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{is}\varepsilon_{it} \xrightarrow{P} \begin{cases} \sigma_{2\varepsilon} & \text{for } t = s \\ 0 & \text{for } t \neq s \end{cases} \quad \text{as } N \to \infty \tag{77}$$

$$\frac{1}{N}\sum_{i=1}^{N}\left(\Delta y_{is} - \varepsilon_{is}\right)\left(\Delta y_{it} - \varepsilon_{it}\right) \xrightarrow{P} 0 \quad \text{as } N \to \infty \tag{78}$$

$$\frac{1}{N}\sum_{i=1}^{N}\left(\Delta y_{is} - \varepsilon_{is}\right)\varepsilon_{it} \xrightarrow{P} 0 \quad \text{as } N \to \infty \tag{79}$$

The result in (77) holds according to Lemma 2. The result in (78) holds since

$$E\left|\frac{1}{N}\sum_{i=1}^{N}\left(\Delta y_{is} - \varepsilon_{is}\right)\left(\Delta y_{it} - \varepsilon_{it}\right)\right| \leq \frac{1}{N}\sum_{i=1}^{N}\sqrt{E\left(\left(\Delta y_{is} - \varepsilon_{is}\right)^2\right)E\left(\left(\Delta y_{it} - \varepsilon_{it}\right)^2\right)} = O\left(N^{-\frac{1}{2}}\right) \tag{80}$$

The inequality results from the triangle inequality and the Cauchy-Schwarz inequality. The equality sign holds by using that $E\left(\left(\Delta y_{is} - \varepsilon_{is}\right)^2\right) = O\left(N^{-\frac{1}{2}}\right)$ according to the result in (72). This shows that $E\left|\frac{1}{N}\sum_{i=1}^{N}\left(\Delta y_{is} - \varepsilon_{is}\right)\left(\Delta y_{it} - \varepsilon_{it}\right)\right| \to 0$ as $N \to \infty$ such that $\frac{1}{N}\sum_{i=1}^{N}\left(\Delta y_{is} - \varepsilon_{is}\right)\left(\Delta y_{it} - \varepsilon_{it}\right) \xrightarrow{P} 0$ as $N \to \infty$. The result in (79) holds since

$$E\left|\frac{1}{N}\sum_{i=1}^{N}\left(\Delta y_{is} - \varepsilon_{is}\right)\varepsilon_{it}\right| \leq \frac{1}{N}\sum_{i=1}^{N}\sqrt{E\left(\left(\Delta y_{is} - \varepsilon_{is}\right)^2\right)E\left(\varepsilon_{it}^2\right)} = O\left(N^{-\frac{1}{4}}\right) \tag{81}$$

Again the inequality results from the triangle inequality and the Cauchy-Schwarz inequality. The equality sign holds by using that $E\left(\varepsilon_{it}^2\right) = O(1)$ and $E\left(\left(\Delta y_{is} - \varepsilon_{is}\right)^2\right) = O\left(N^{-\frac{1}{2}}\right)$ according to the result in (72). This shows that $E\left|\frac{1}{N}\sum_{i=1}^{N}\left(\Delta y_{is} - \varepsilon_{is}\right)\varepsilon_{it}\right| \to 0$ as $N \to \infty$ such that $\frac{1}{N}\sum_{i=1}^{N}\left(\Delta y_{is} - \varepsilon_{is}\right)\varepsilon_{it} \xrightarrow{P} 0$ as $N \to \infty$. Altogether, this proves the result in (b).

(c) We prove the result by showing that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}Q_i'\varepsilon_i \xrightarrow{w} N\left(0, \sigma_{4\varepsilon}I_m\right) \quad \text{as } N \to \infty \tag{82}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(Z_{i1} - Q_i\right)' v_i \xrightarrow{P} 0 \quad \text{as } N \to \infty \tag{83}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}Q_i'\left(v_i - \varepsilon_i\right) = (1-\rho)\frac{1}{\sqrt{N}}\sum_{i=1}^{N}Q_i'\iota_{T-1}\alpha_i \xrightarrow{P} 0 \quad \text{as } N \to \infty \tag{84}$$

where $\varepsilon_i = \left(\varepsilon_{i2}, ..., \varepsilon_{iT}\right)$ and the $(T-1) \times m$ matrix Q_i is defined as

$$Q_i = \begin{bmatrix} \varepsilon_{i1} & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \varepsilon_{i1} & \varepsilon_{i2} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & \varepsilon_{i1} & \cdots & \varepsilon_{iT-1} \end{bmatrix} \tag{85}$$

The result in (82) follows by the Liapounov Central Limit Theorem as ε_{it} is independent of ε_{is} for $s \neq t$. As $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(Z_{i1} - Q_i\right)' v_i$ has elements of the form $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\Delta y_{is} - \varepsilon_{is}\right)v_{it}$ for $t = 2, ..., T$ and $s = 1, ..., t-1$, the result in (83) holds since

$$E\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\Delta y_{is} - \varepsilon_{is}\right)v_{it}\right|^2 = \frac{1}{N}\sum_{i=1}^{N}E\left(\left(\Delta y_{is} - \varepsilon_{is}\right)^2\right)E\left(v_{it}^2\right) = O\left(N^{-\frac{1}{2}}\right) \tag{86}$$

The first equality sign results from $(\Delta y_{is} - \varepsilon_{is})$ and v_{it} being independent with means zero such that $(\Delta y_{is} - \varepsilon_{is}) v_{it}$ is independent across i with mean zero. The second equality sign follows by using that $E\left((\Delta y_{is} - \varepsilon_{is})^2\right) = O\left(N^{-\frac{1}{2}}\right)$ and $E\left(v_{it}^2\right) = O(1)$ according to (72) and (73). This shows that $E\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\Delta y_{is} - \varepsilon_{is}) v_{it}\right|^2 \to 0$ as $N \to \infty$ such that $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\Delta y_{is} - \varepsilon_{is}) v_{it} \xrightarrow{P} 0$ as $N \to \infty$. The result in (84) follows by using that $\frac{1}{\sqrt{N}}\sum_{i=1}^{N} Q_i' \iota_{T-1}\alpha_i \xrightarrow{w} N\left(0, \sigma_{2\varepsilon}\sigma_\alpha^2\right)$ as $N \to \infty$ by the Liapounov Central Limit Theorem and that $(1-\rho) = c/\sqrt{N} \to 0$ as $N \to \infty$. Altogether, we have obtained the desired limits and the result in (c) is proved.

(d1) and (d2) These results follow directly by the results already obtained. $\square$

Lemma 4 *Under Assumption 1, 2, 3 and the local-to-unity sequence for ρ given by $\rho = 1 - c/\sqrt{N}$ for $c \geq 0$, the following result holds*

$$\frac{1}{N}\sum_{i=1}^{N} Z_{i1}'\hat{v}_i\hat{v}_i' Z_{i1} \xrightarrow{P} \sigma_{4\varepsilon} I_m \quad \text{as } N \to \infty \tag{a}$$

where $\hat{v}_i = y_i - \tilde{\rho}_I y_{i,-1}$ and $\tilde{\rho}_I$ denotes any one-step $\sqrt{N}$-consistent estimator of ρ, for example the one defined in (53).

Proof of Lemma 4:

Inserting the expression for $\hat{v}_i$ given by $(\rho - \tilde{\rho}_I) y_{i,-1} + v_i$ yields

$$\begin{aligned} & \frac{1}{N}\sum_{i=1}^{N} Z_{i1}'\hat{v}_i\hat{v}_i' Z_{i1} \\ = \; & \frac{1}{N}\sum_{i=1}^{N} Z_{i1}' v_i v_i' Z_{i1} + (\rho - \tilde{\rho}_I)^2 \frac{1}{N}\sum_{i=1}^{N} Z_{i1}' y_{i,-1} y_{i,-1}' Z_{i1} + 2(\rho - \tilde{\rho}_I)\frac{1}{N}\sum_{i=1}^{N} Z_{i1}' v_i y_{i,-1}' Z_{i1} \end{aligned} \tag{87}$$

As $\tilde{\rho}_I$ is $\sqrt{N}$-consistent we have that $\sqrt{N}(\tilde{\rho}_I - \rho) = O_P(1)$ and therefore we prove the result by showing that

$$\frac{1}{N}\sum_{i=1}^{N} Z_{i1}' v_i v_i' Z_{i1} \xrightarrow{P} \sigma_{4\varepsilon} I_m \quad \text{as } N \to \infty \tag{88}$$

$$\frac{1}{N^2}\sum_{i=1}^{N} Z_{i1}' y_{i,-1} y_{i,-1}' Z_{i1} \xrightarrow{P} 0 \quad \text{as } N \to \infty \tag{89}$$

$$\frac{1}{N^{\frac{3}{2}}}\sum_{i=1}^{N} Z_{i1}' v_i y_{i,-1}' Z_{i1} \xrightarrow{P} 0 \quad \text{as } N \to \infty \tag{90}$$

The results below are used in the following

$$E\left(y_{it}^4\right)^{\frac{1}{4}} \leq E\left(\alpha_i^4\right)^{\frac{1}{4}} + \rho^t \tau(\rho)^{\frac{1}{2}} E\left(\varepsilon_{i0}^4\right)^{\frac{1}{4}} + \sum_{j=1}^{t} \rho^{t-j} E\left(\varepsilon_{ij}^4\right)^{\frac{1}{4}} = O\left(N^{\frac{1}{4}}\right) \tag{91}$$

$$E\left(\Delta y_{it}^4\right)^{\frac{1}{4}} \leq (\rho-1)\rho^{t-1}\tau(\rho)^{\frac{1}{2}} E\left(\varepsilon_{i0}^4\right)^{\frac{1}{4}} + (\rho-1)\sum_{j=1}^{t-1} \rho^{t-1-j} E\left(\varepsilon_{ij}^4\right)^{\frac{1}{4}} + E\left(\varepsilon_{it}^4\right)^{\frac{1}{4}} = O(1) \tag{92}$$

$$E\left(\left(\Delta y_{it} - \varepsilon_{it}\right)^4\right)^{\frac{1}{4}} \leq (\rho-1)\rho^{t-1}\tau(\rho)^{\frac{1}{2}} E\left(\varepsilon_{i0}^4\right)^{\frac{1}{4}} + (\rho-1)\sum_{j=1}^{t-1} \rho^{t-1-j} E\left(\varepsilon_{ij}^4\right)^{\frac{1}{4}} = O\left(N^{-\frac{1}{4}}\right) \tag{93}$$

$$E\left(v_{it}^4\right)^{\frac{1}{4}} \leq (1-\rho) E\left(\alpha_i^4\right)^{\frac{1}{4}} + E\left(\varepsilon_{it}^4\right)^{\frac{1}{4}} = O(1) \tag{94}$$

where the inequalities result from Minkowski's inequality and the equalities hold according to Assumption 3 and the results in Lemma 1.

We prove the result in (88) by showing that

$$\frac{1}{N}\sum_{i=1}^{N} Q_i' v_i v_i' Q_i \xrightarrow{P} \sigma_{4\varepsilon} I_m \quad \text{as } N \to \infty \tag{95}$$

$$\frac{1}{N}\sum_{i=1}^{N} \left(Z_{i1} - Q_i\right)' v_i v_i' Z_{i1} \xrightarrow{P} 0 \quad \text{as } N \to \infty \tag{96}$$

$$\frac{1}{N}\sum_{i=1}^{N} \left(Z_{i1} - Q_i\right)' v_i v_i' Q_i \xrightarrow{P} 0 \quad \text{as } N \to \infty \tag{97}$$

where Q_i is defined in (85). $\frac{1}{N}\sum_{i=1}^{N} Q_i' v_i v_i' Q_i$ has elements of the form $\frac{1}{N}\sum_{i=1}^{N} \varepsilon_{is}\varepsilon_{il} v_{it} v_{ik}$ for $t, k = 2, ..., T$, $s = 1, ..., t-1$ and $l = 1, ..., k-1$ where the diagonal elements correspond to $t = k$ and $s = l$. We have

$$\begin{aligned} & \frac{1}{N}\sum_{i=1}^{N} \varepsilon_{is}\varepsilon_{il} v_{it} v_{ik} \\ = & \frac{1}{N}\sum_{i=1}^{N} \varepsilon_{is}\varepsilon_{il}\varepsilon_{it}\varepsilon_{ik} + (1-\rho)^2 \frac{1}{N}\sum_{i=1}^{N} \varepsilon_{is}\varepsilon_{il}\alpha_i^2 + (1-\rho)\frac{1}{N}\sum_{i=1}^{N} \varepsilon_{is}\varepsilon_{il}\varepsilon_{ik}\alpha_i + (1-\rho)\frac{1}{N}\sum_{i=1}^{N} \varepsilon_{is}\varepsilon_{il}\varepsilon_{it}\alpha_i \end{aligned}$$

Using this the result in (95) now follows since

$$\frac{1}{N}\sum_{i=1}^{N} \varepsilon_{is}\varepsilon_{il}\varepsilon_{it}\varepsilon_{ik} \xrightarrow{P} \begin{cases} \sigma_{4\varepsilon} & s = l, t = k \\ 0 & s \neq l \end{cases} \quad \text{as } N \to \infty \tag{98}$$

$$\frac{1}{N}\sum_{i=1}^{N} \varepsilon_{is}\varepsilon_{il}\alpha_i^2 \xrightarrow{P} \begin{cases} \sigma_{2\varepsilon}\sigma_\alpha^2 & s = l \\ 0 & s \neq l \end{cases} \quad \text{as } N \to \infty \tag{99}$$

$$\frac{1}{N}\sum_{i=1}^{N} \varepsilon_{is}\varepsilon_{il}\varepsilon_{ik}\alpha_k \xrightarrow{P} 0 \quad \text{as } N \to \infty \quad \text{for all } k = 2, ..., T \tag{100}$$

$$(1-\rho)^2 = \frac{c^2}{N} \to 0 \quad \text{as } N \to \infty \tag{101}$$

where the first three results follow by Markow's Law of Large Numbers. To show the result in (96) we note that $\frac{1}{N}\sum_{i=1}^{N} \left(Z_{i1} - Q_i\right)' v_i v_i' Z_{i1}$ has elements of the form $\frac{1}{N}\sum_{i=1}^{N} \left(\Delta y_{is} - \varepsilon_{is}\right) \Delta y_{il} v_{it} v_{ik}$ where

$t, k = 2, ..., T,\ s = 1, ..., t-1$ and $l = 1, ..., k-1$. We have that

$$\begin{aligned} E\left|\frac{1}{N}\sum_{i=1}^{N}\left(\Delta y_{is}-\varepsilon_{is}\right)\Delta y_{il}v_{it}v_{ik}\right| &\leq \frac{1}{N}\sum_{i=1}^{N}\left(E\left(\left(\Delta y_{is}-\varepsilon_{is}\right)^{4}\right)E\left(\Delta y_{il}^{4}\right)E\left(v_{it}^{4}\right)E\left(v_{ik}^{4}\right)\right)^{\frac{1}{4}} \\ &= O\left(N^{-\frac{1}{4}}\right) \end{aligned} \tag{102}$$

where the inequality results from the triangle inequality and the Cauchy-Schwarz inequality and the equality sign holds by using the results in (91)-(94). This shows that $E\left|\frac{1}{N}\sum_{i=1}^{N}\left(Z_{i1}-Q_{i}\right)' v_{i}v_{i}'Z_{i1}\right| \to 0$ as $N \to \infty$ which proves (96). To show the result in (97) we note that $\frac{1}{N}\sum_{i=1}^{N}\left(Z_{i1}-Q_{i}\right)' v_{i}v_{i}'Q_{i}$ has elements of the form $\frac{1}{N}\sum_{i=1}^{N}\left(\Delta y_{is}-\varepsilon_{is}\right)\Delta y_{il}v_{it}\varepsilon_{ik}$ where $t, k = 2, ..., T,\ s = 1, ..., t-1$ and $l = 1, ..., k-1$. By using similar arguments as above we have that

$$\begin{aligned} E\left|\frac{1}{N}\sum_{i=1}^{N}\left(\Delta y_{is}-\varepsilon_{is}\right)\Delta y_{il}v_{it}\varepsilon_{ik}\right| &\leq \frac{1}{N}\sum_{i=1}^{N}\left(E\left(\left(\Delta y_{is}-\varepsilon_{is}\right)^{4}\right)E\left(\Delta y_{il}^{4}\right)E\left(v_{it}^{4}\right)E\left(\varepsilon_{ik}^{4}\right)\right)^{\frac{1}{4}} \\ &= O\left(N^{-\frac{1}{4}}\right) \end{aligned} \tag{103}$$

such that $E\left|\frac{1}{N}\sum_{i=1}^{N}\left(Z_{i1}-Q_{i}\right)' v_{i}v_{i}'Q_{i}\right| \to 0$ as $N \to \infty$ which proves (97). Altogether, this proves the result in (88).

To show the result in (89) we note that the elements in $\frac{1}{N^2}\sum_{i=1}^{N} Z_{i1}'y_{i,-1}y_{i,-1}'Z_{i1}$ are on the form $\frac{1}{N^2}\sum_{i=1}^{N} y_{it}y_{ik}\Delta y_{is}\Delta y_{il}$ for $t, k = 1, ..., T-1$ and $s = 1, ..., t$ and $l = 1, ..., k$. We have that

$$E\left|\frac{1}{N^2}\sum_{i=1}^{N} y_{it}y_{ik}\Delta y_{is}\Delta y_{il}\right| \leq \frac{1}{N^2}\sum_{i=1}^{N}\left(E\left(y_{it}^{4}\right)E\left(y_{ik}^{4}\right)E\left(\Delta y_{is}^{4}\right)E\left(\Delta y_{il}^{4}\right)\right)^{\frac{1}{4}} = \frac{1}{N}O\left(N^{\frac{1}{2}}\right) \tag{104}$$

where the inequalities follow by the triangle inequality and the Cauchy-Schwarz inequality and the equality sign holds by the results in (91)-(94). This shows that $E\left|\frac{1}{N^2}\sum_{i=1}^{N} Z_{i1}'y_{i,-1}y_{i,-1}'Z_{i1}\right| \to 0$ as $N \to \infty$ which proves (89). To show (90) we note that the elements in $\frac{1}{N^{\frac{3}{2}}}\sum_{i=1}^{N} Z_{i1}'v_{i}y_{i,-1}'Z_{i1}$ are on the form $\frac{1}{N^{\frac{3}{2}}}\sum_{i=1}^{N} y_{it}v_{ik}\Delta y_{is}\Delta y_{il}$ where $t = 1, ..., T-1,\ k = 2, ..., T,\ s = 1, ..., t$ and $l = 1, ..., k-1$. Using similar arguments as above the following holds

$$E\left|\frac{1}{N^{\frac{3}{2}}}\sum_{i=1}^{N} y_{it}v_{ik}\Delta y_{is}\Delta y_{il}\right| \leq \frac{1}{N^{\frac{3}{2}}}\sum_{i=1}^{N}\left(E\left(y_{it}^{4}\right)E\left(v_{ik}^{4}\right)E\left(\Delta y_{is}^{4}\right)E\left(\Delta y_{il}^{4}\right)\right)^{\frac{1}{4}} = \frac{1}{N^{\frac{1}{2}}}O\left(N^{\frac{1}{4}}\right) \tag{105}$$

such that $E\left|\frac{1}{N^2}\sum_{i=1}^{N} Z_{i1}'v_{i}y_{i,-1}'Z_{i1}\right| \to 0$ as $N \to \infty$ which proves (90). Altogether we have obtained the desired limits and the result is proved. $\square$

Proof of Proposition 1 and 2:

The optimal two-step estimator $\hat{\rho}_I$ corresponding to (16) can be expressed as

$$\hat{\rho}_I = \rho + \left(\sum_{i=1}^{N} y_{i,-1}'Z_{i1}\left(\sum_{i=1}^{N} Z_{i1}'\hat{v}_i\hat{v}_i'Z_{i1}\right)^{-1}\sum_{i=1}^{N} Z_{i1}'y_{i,-1}\right)^{-1}\sum_{i=1}^{N} y_{i,-1}'Z_{i1}\left(\sum_{i=1}^{N} Z_{i1}'\hat{v}_i\hat{v}_i'Z_{i1}\right)^{-1}\sum_{i=1}^{N} Z_{i1}'v_i$$

The results in Proposition 1 now follow immediately from the results in Lemma 3 and 4. These results also show that the one-step estimator $\tilde{\rho}_I$ is asymptotically equivalent to the optimal two-step estimator

$\hat{\rho}_I$. The t-statistic defined in (20) can be expressed as

$$t_I = \left(\frac{1}{N} \sum_{i=1}^{N} y_{i,-1}' Z_{i1} \left(\frac{1}{N} \sum_{i=1}^{N} Z_{i1}' \hat{v}_i \hat{v}_i' Z_{i1} \right)^{-1} \frac{1}{N} \sum_{i=1}^{N} Z_{i1}' y_{i,-1} \right)^{\frac{1}{2}} \left(\sqrt{N} \left(\hat{\rho}_I - \rho \right) - c \right) \tag{106}$$

and the results in Proposition 2 now follow immediately from the results in Proposition 1 together with the results in Lemma 3 and 4. □

Proof of Proposition 3:

The results follow directly by Lemma 5 given below. □

Lemma 5 *Under Assumption 1, 2, 3 and the local-to-unity sequence for ρ given by $\rho = 1 - c/\sqrt{N}$ for $c \geq 0$, the following results hold*

$$(i) \text{ and } c \geq 0: \qquad \frac{1}{\sqrt{N}} \sum_{i=1}^{N} Z_{i1}' \Delta y_i \xrightarrow{w} N \left(-c\sigma_{2\varepsilon} \iota_m, \sigma_{4\varepsilon} I_m \right) \quad \text{as } N \to \infty \tag{a1}$$

$$(ii) \text{ and } c > 0: \qquad \frac{1}{\sqrt{N}} \sum_{i=1}^{N} Z_{i1}' \Delta y_i \xrightarrow{w} N \left(-\frac{1}{2} c\sigma_{2\varepsilon} \iota_m, \sigma_{4\varepsilon} I_m \right) \quad \text{as } N \to \infty \tag{a2}$$

$$\frac{1}{N} \sum_{i=1}^{N} Z_{i1}' \Delta y_i \Delta y_i' Z_{i1} \xrightarrow{P} \sigma_{4\varepsilon} I_m \quad \text{as } N \to \infty \tag{b}$$

Proof of Lemma 5:

(a1) and (a2) Using that $\Delta y_i = (\rho - 1) y_{i,-1} + v_i$ we have

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} Z_{i1}' \Delta y_i = (\rho - 1) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} Z_{i1}' y_{i,-1} + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} Z_{i1}' v_i = -c \frac{1}{N} \sum_{i=1}^{N} Z_{i1}' y_{i,-1} + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} Z_{i1}' v_i \tag{107}$$

and the results now follow by (a1), (a2) and (c) in Lemma 3.

(b) Again using the expression for Δy_i given above we have

$$\begin{aligned}
& \frac{1}{N} \sum_{i=1}^{N} Z_{i1}' \Delta y_i \Delta y_i' Z_{i1} \\
= & \frac{1}{N} \sum_{i=1}^{N} Z_{i1}' v_i v_i' Z_{i1} + 2 (\rho - 1) \frac{1}{N} \sum_{i=1}^{N} Z_{i1}' y_{i,-1} v_i' Z_{i1} + (\rho - 1)^2 \frac{1}{N} \sum_{i=1}^{N} Z_{i1}' y_{i,-1} y_{i,-1}' Z_{i1} \\
= & \frac{1}{N} \sum_{i=1}^{N} Z_{i1}' v_i v_i' Z_{i1} - 2c \frac{1}{N^{\frac{3}{2}}} \sum_{i=1}^{N} Z_{i1}' y_{i,-1} v_i' Z_{i1} + c^2 \frac{1}{N^2} \sum_{i=1}^{N} Z_{i1}' y_{i,-1} y_{i,-1}' Z_{i1}
\end{aligned} \tag{108}$$

and the result now follows by the results in (88), (89) and (90). □

A.3 Proof of the propositions in Section 3.2

For repetition the regression model is the following

$$\begin{aligned}
\Delta y_{it} &= \rho \Delta y_{it-1} + \Delta v_{it} \\
\Delta v_{it} &= \Delta \varepsilon_{it}
\end{aligned} \qquad \text{for } i = 1, ..., N \text{ and } t = 2, ..., T \tag{109}$$

Using stacked notation the regressor is $\Delta y_{i,-1} = (\Delta y_{i1}, ..., \Delta y_{iT-1})'$, the regression error is $\Delta v_i = (\Delta v_{i2}, ..., \Delta v_{iT})'$ and the $(T-1) \times m$ matrix of instruments is

$$Z_{i2} = \begin{bmatrix} y_{i0} & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & y_{i0} & y_{i1} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & y_{i0} & \cdots & y_{iT-2} \end{bmatrix} \tag{110}$$

Let the $m \times 1$ vectors X_{1N} and X_{2N} be defined in the following way

$$X_{1N} = \sum_{i=1}^{N} Z_{i2}' \Delta v_i \tag{111}$$

$$X_{2N} = \sum_{i=1}^{N} Z_{i2}' \Delta y_{i,-1} \tag{112}$$

The standard IV estimator $\hat{\rho}_{II}$ corresponding to the instrumental variables regression above can then be expressed in the following way

$$\hat{\rho}_{II} = \rho + \left(X_{2N}' \left(\sum_{i=1}^{N} Z_{i2}' Z_{i2} \right)^{-1} X_{2N} \right)^{-1} \sum_{i=1}^{N} X_{2N}' \left(\sum_{i=1}^{N} Z_{i2}' Z_{i2} \right)^{-1} X_{1N} \tag{113}$$

Proof of Proposition 4:

The result follows directly by Lemma 6 below. $\square$

Lemma 6 *Under Assumption 1, 2(i), 3 and the local-to-unity sequence for ρ given by $\rho = 1 - c/\sqrt{N}$ for $c \geq 0$, the following results hold*

$$\frac{1}{N} \sum_{i=1}^{N} Z_{i2}' Z_{i2} \xrightarrow{P} \tilde{\Sigma}_{11} \quad \text{as } N \to \infty \tag{a}$$

The $m \times m$ matrix $\tilde{\Sigma}_{11}$ is defined as

$$\tilde{\Sigma}_{11} = \begin{bmatrix} \tilde{P}_1 & 0 & 0 & \cdots & 0 \\ 0 & \tilde{P}_2 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \tilde{P}_{T-2} & 0 \\ 0 & \cdots & 0 & 0 & \tilde{P}_{T-1} \end{bmatrix} \tag{114}$$

where $\tilde{P}_k$ for $k = 1, ..., T-1$ is $k \times k$ matrix where element i, j is the following

$$P_k(i,j) = \mathbf{1}_{\{c>0\}} \sigma_\alpha^2 + (\min\{i,j\} - 1 + \tau)\, \sigma_{2\varepsilon} \tag{115}$$

In addition the following holds

$$\frac{1}{\sqrt{N}} \begin{bmatrix} X_{1N} \\ X_{2N} \end{bmatrix} \xrightarrow{w} N\left(\begin{bmatrix} 0 \\ -cq \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{bmatrix} \right) \quad \text{as } N \to \infty \tag{b}$$

The $m \times 1$ vector q is defined as $q = \sigma_{2\varepsilon} \left(q_1', q_2', ..., q_{T-1}'\right)'$ where q_k for $k = 1, ..., T-1$ is a $k \times 1$ vector where element i is equal to $\tau + i - 1$. The $m \times m$ matrices Σ_{11}, Σ_{12} and Σ_{22} are defined as

$$\Sigma_{11} = \begin{bmatrix} 2P_{11} & -P_{12} & 0 & \cdots & 0 \\ -P_{21} & 2P_{22} & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 2P_{T-2,T-2} & -P_{T-2,T-1} \\ 0 & \cdots & 0 & -P_{T-1,T-2} & 2P_{T-1,T-1} \end{bmatrix} \tag{116}$$

$$\Sigma_{12} = \begin{bmatrix} -P_{11} & P_{12} & 0 & \cdots & 0 \\ 0 & -P_{22} & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & -P_{T-2,T-2} & P_{T-2,T-1} \\ 0 & \cdots & 0 & 0 & -P_{T-1,T-1} \end{bmatrix} \tag{117}$$

$$\Sigma_{22} = \begin{bmatrix} P_{11} & 0 & 0 & \cdots & 0 \\ 0 & P_{22} & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & P_{T-2,T-2} & 0 \\ 0 & \cdots & 0 & 0 & P_{T-1,T-1} \end{bmatrix} \tag{118}$$

where P_{kl} for $k, l = 1, ..., T-1$ is a $k \times l$ matrix where element i, j is the following

$$P_{kl}(i,j) = \sigma_{2\varepsilon} \left(\mathbf{1}_{\{c>0\}} \sigma_\alpha^2 + (\min\{i,j\} - 1 + \tau) \frac{\sigma_{4\varepsilon}}{\sigma_{2\varepsilon}} \right) \tag{119}$$

Proof of Lemma 6:

(a) $\frac{1}{N} \sum_{i=1}^N Z_{i2}' Z_{i2}$ has non-zero elements of the form $\frac{1}{N} \sum_{i=1}^N y_{it} y_{is}$ for $t = 0, ..., T-2$ and $s = 0, ..., t$. We prove the result by showing that

$$\frac{1}{N} \sum_{i=1}^N y_{it} y_{is} \xrightarrow{P} \mathbf{1}_{\{c>0\}} \sigma_\alpha^2 + (\min\{s,t\} - 1 + \tau) \sigma_{2\varepsilon} \quad \text{as } N \to \infty \tag{120}$$

This follows by Lemma 2 and by using that $\frac{1}{N} \sum_{i=1}^N \varepsilon_{is} \varepsilon_{it} \xrightarrow{P} 0$ as $N \to \infty$ for $s \neq t$, $\frac{1}{N} \sum_{i=1}^N \alpha_i \varepsilon_{it} \xrightarrow{P} 0$ as $N \to \infty$ and $\rho^k \to 1$ as $N \to \infty$.

(b) Using the expressions for y_{it} and Δy_{it} in (45) and (46), we first of all note that when $\tau(\rho) = \tau$ then according to Assumption 3

$$E|y_{is} \Delta y_{it}|^{2+\delta/2} \leq K_1 < \infty \quad \text{for all } i = 1, ..., N \tag{121}$$

$$E|y_{is} \Delta v_{it}|^{2+\delta/2} \leq K_2 < \infty \quad \text{for all } i = 1, ..., N \tag{122}$$

for all $s = 0, ..., T-2$ and $t = 2, ..., T$. This means that the Liapounov Central Limit Theorem can be applied to the sequence $\frac{1}{\sqrt{N}} \sum_{i=1}^N (X_{1N}', X_{2N}')'$. We have that

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N E(X_{1N}) = 0 \tag{123}$$

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N E(X_{2N}) \to -cq \quad \text{as } N \to \infty \tag{124}$$

With respect to the covariance matrix we use the following results. By using the expressions for y_{it} and Δy_{it} in (45) and (46) and that Δv_{it} is independent of y_{is} for $t = 2, ..., T$ and $s = 0, ..., t-2$ we find that $t, k = 2, ..., T$ and $s = 0, ..., t-2, l = 0, ..., k-2$

$$\frac{1}{N}\sum_{i=1}^{N} E\left(y_{is}y_{it}\right)\sigma_{i\varepsilon}^2 \to \sigma_{2\varepsilon}\left(\mathbf{1}_{\{c>0\}}\sigma_\alpha^2 + \left(\min\{s,t\} - 1 + \tau\right)\frac{\sigma_{4\varepsilon}}{\sigma_{2\varepsilon}}\right) \quad \text{as } N \to \infty \tag{125}$$

$$\frac{1}{N}\sum_{i=1}^{N} E\left(y_{is}y_{il}\Delta y_{it}^2\right) \to \sigma_{2\varepsilon}\left(\mathbf{1}_{\{c>0\}}\sigma_\alpha^2 + \left(\min\{s,l\} - 1 + \tau\right)\frac{\sigma_{4\varepsilon}}{\sigma_{2\varepsilon}}\right) \quad \text{as } N \to \infty \tag{126}$$

$$\frac{1}{N}\sum_{i=1}^{N} E\left(y_{is}y_{il}\Delta y_{it}\Delta y_{ik}\right) \to 0 \quad \text{as } N \to \infty \quad \text{for } t \neq k \tag{127}$$

$$\frac{1}{N}\sum_{i=1}^{N} E\left(y_{is}y_{il}\Delta v_{it}\Delta y_{it-1}\right) \to -\sigma_{2\varepsilon}\left(\mathbf{1}_{\{c>0\}}\sigma_\alpha^2 + \left(\min\{s,l\} - 1 + \tau\right)\frac{\sigma_{4\varepsilon}}{\sigma_{2\varepsilon}}\right) \quad \text{as } N \to \infty \tag{128}$$

$$\frac{1}{N}\sum_{i=1}^{N} E\left(y_{is}y_{il}\Delta v_{it}\Delta y_{it}\right) \to \sigma_{2\varepsilon}\left(\mathbf{1}_{\{c>0\}}\sigma_\alpha^2 + \left(\min\{s,l\} - 1 + \tau\right)\frac{\sigma_{4\varepsilon}}{\sigma_{2\varepsilon}}\right) \quad \text{as } N \to \infty \tag{129}$$

The matrix $Z_{i2}'\Delta v_i\Delta v_i' Z_{i2}$ has elements of the form $y_{is}y_{il}\Delta v_{it}\Delta v_{ik}$ where $t, k = 2, ..., T$ and $s = 0, ..., t-2, l = 0, ..., k-2$. Using this we have

$$E\left(y_{is}y_{il}\Delta v_{it}\Delta v_{ik}\right) = \begin{cases} 2E\left(y_{is}^2\right)\sigma_{i\varepsilon}^2 & \text{for } t = k, s = l \\ -E\left(y_{is}y_{il}\right)\sigma_{i\varepsilon}^2 & \text{for } |t-k| = 1 \\ 0 & \text{for } |t-k| > 1 \end{cases} \tag{130}$$

such that

$$\frac{1}{N}\sum_{i=1}^{N} E\left(Z_{i2}'\Delta v_i\Delta v_i' Z_{i2}\right) \to \Sigma_{11} \quad \text{as } N \to \infty \tag{131}$$

The matrix $Z_{i2}'\Delta y_{i,-1}\Delta y_{i,-1}' Z_{i2}$ has elements of the form $y_{is}y_{il}\Delta y_{it}\Delta y_{ik}$ where $t, k = 1, ..., T-1$ and $s = 0, ..., t-1, l = 0, ..., k-1$. Using this we have

$$E\left(y_{is}y_{il}\Delta y_{it}\Delta y_{ik}\right) = \begin{cases} E\left(y_{is}y_{il}\Delta y_{it}^2\right) & \text{for } t = k \\ E\left(y_{is}y_{il}\Delta y_{it}\Delta y_{ik}\right) & \text{for } t \neq k \end{cases} \tag{132}$$

such that

$$\frac{1}{N}\sum_{i=1}^{N} E\left(Z_{i2}'\Delta y_{i,-1}\Delta y_{i,-1}' Z_{i2}\right) \to \Sigma_{22} \quad \text{as } N \to \infty \tag{133}$$

The matrix $Z_{i2}'\Delta v_i\Delta y_{i,-1}' Z_{i2}$ has elements of the form $y_{is}y_{il}\Delta v_{it+1}\Delta y_{ik}$ where $t, k = 1, ..., T-1$ and $s = 0, ..., t-1, l = 0, ..., k-1$. Using this we have

$$E\left(y_{is}y_{il}\Delta v_{it+1}\Delta y_{ik}\right) = \begin{cases} E\left(y_{is}y_{il}\right)E\left(\Delta v_{it+1}\Delta y_{it}\right) & \text{for } t = k \\ E\left(y_{is}y_{il}\right)E\left(\Delta v_{it}\Delta y_{it}\right) & \text{for } t = k-1 \\ 0 & \text{otherwise} \end{cases} \tag{134}$$

such that

$$\frac{1}{N}\sum_{i=1}^{N} E\left(Z_{i2}'\Delta v_i\Delta y_{i,-1}' Z_{i2}\right) \to \Sigma_{12} \quad \text{as } N \to \infty \tag{135}$$

Altogether, this proves the result in (b). $\square$

Lemma 7 *Under Assumption 1, 2(ii), 3 and the local-to-unity sequence for ρ given by $\rho = 1 - \tilde{c}/N$ for $\tilde{c} > 0$, the following results hold*

$$\frac{1}{N}\sum_{i=1}^{N} Z_{i2}'\Delta y_{i,-1} \xrightarrow{w} N\left(-\frac{1}{2}\sigma_{2\varepsilon}\iota_m, \frac{1}{2\tilde{c}}\sigma_{4\varepsilon}\Gamma\right) \quad \textit{as } N \to \infty \tag{a}$$

$$\frac{1}{N^2}\sum_{i=1}^{N} Z_{i2}'\Delta y_{i,-1}\Delta y_{i,-1}' Z_{i2} \xrightarrow{P} \frac{1}{2\tilde{c}}\sigma_{4\varepsilon}\Gamma \quad \textit{as } N \to \infty \tag{b}$$

The $m \times m$ matrix Γ is defined as

$$\Gamma = \begin{bmatrix} \Gamma_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \Gamma_{T-1} \end{bmatrix} \tag{136}$$

where Γ_k for $k = 1, ..., T-1$ is a symmetric $k \times k$ matrix with all elements equal to one.

Proof of Lemma 7:

(a) Using the expressions for Z_{i2} and $\Delta y_{i,-1}$ we have

$$Z_{i2}'\Delta y_{i,-1} = Q_i'\left(R_{1i} + R_{2i}\right) + \left(Z_{i2} - Q_i\right)'\Delta y_{i,-1} + Q_i'\left(\Delta y_{i,-1} - R_{1i} - R_{2i}\right) \tag{137}$$

where

$$Q_i = \tau(\rho)^{\frac{1}{2}}\varepsilon_{i0}\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \rho & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & \cdots & \rho^{T-2} \end{bmatrix} \tag{138}$$

$$R_{1i} = (\rho - 1)\tau(\rho)^{\frac{1}{2}}\varepsilon_{i0}\begin{bmatrix} 1 \\ \rho \\ \vdots \\ \rho^{T-2} \end{bmatrix} \qquad R_{2i} = \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{iT-1} \end{bmatrix} \tag{139}$$

We prove the result by showing that

$$\frac{1}{N}\sum_{i=1}^{N} Q_i'\left(R_{1i} + R_{2i}\right) \xrightarrow{w} N\left(-\frac{1}{2}\sigma_{2\varepsilon}\iota_m, \frac{1}{2\tilde{c}}\sigma_{4\varepsilon}\Gamma\right) \quad \text{as } N \to \infty \tag{140}$$

$$\frac{1}{N}\sum_{i=1}^{N} \left(Z_{i2} - Q_i\right)'\Delta y_{i,-1} \xrightarrow{P} 0 \quad \text{as } N \to \infty \tag{141}$$

$$\frac{1}{N}\sum_{i=1}^{N} Q_i'\left(\Delta y_{i,-1} - R_{1i} - R_{2i}\right) \xrightarrow{P} 0 \quad \text{as } N \to \infty \tag{142}$$

The result in (140) is proved by showing that

$$\frac{1}{N}\sum_{i=1}^{N} Q_i' R_{1i} \xrightarrow{P} -\frac{1}{2}\sigma_{2\varepsilon}\iota_m \quad \text{as } N \to \infty \tag{143}$$

$$\frac{1}{N}\sum_{i=1}^{N} Q_i' R_{2i} \xrightarrow{w} N\left(0, \frac{1}{2\tilde{c}}\sigma_{4\varepsilon}\Gamma\right) \quad \text{as } N \to \infty \tag{144}$$

The result in (143) holds since

$$\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i0}^{2}\xrightarrow{P}\sigma_{2\varepsilon}\quad \text{as } N\to\infty \tag{145}$$

$$(\rho-1)\,\tau(\rho)\to-\frac{1}{2}\quad \text{as } N\to\infty \tag{146}$$

$$\rho^{t}\to 1\quad \text{as } N\to\infty \tag{147}$$

and the result in (144) holds since

$$\tau(\rho)^{-\frac{1}{2}}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}Q_i'R_{2i}\xrightarrow{w}N(0,\sigma_{4\varepsilon}\Gamma)\quad \text{as } N\to\infty \tag{148}$$

$$\frac{1}{\sqrt{N}}\tau(\rho)^{\frac{1}{2}}\to\frac{1}{(2\tilde{c})^{\frac{1}{2}}}\quad \text{as } N\to\infty \tag{149}$$

Altogether, this shows the result in (140). To show the result in (141) we note that $(Z_{i2}-Q_i)'\,\Delta y_{i,-1}$ has elements of the form $\left(y_{is}-\tau(\rho)^{\frac{1}{2}}\rho^{s}\varepsilon_{i0}\right)\Delta y_{it}$ for $t=1,...,T-1$ and $s=1,...,t-1$. We have

$$\begin{aligned} &E\left|\frac{1}{N}\sum_{i=1}^{N}\left(y_{is}-\tau(\rho)^{\frac{1}{2}}\rho^{s}\varepsilon_{i0}\right)(\Delta y_{it}-\varepsilon_{it})\right| \\ \leq\ & \frac{1}{N}\sum_{i=1}^{N}\sqrt{E\left(\left(y_{is}-\tau(\rho)^{\frac{1}{2}}\rho^{s}\varepsilon_{i0}\right)^{2}\right)E\left((\Delta y_{it}-\varepsilon_{it})^{2}\right)}=O\left(N^{-\frac{1}{2}}\right) \end{aligned} \tag{150}$$

where the inequality results from the triangle equality and the Cauchy-Schwarz inequality and the equality sign holds by using that $E\left(\left(y_{is}-\tau(\rho)^{\frac{1}{2}}\rho^{s}\varepsilon_{i0}\right)^{2}\right)=O(1)$ and $E\left((\Delta y_{it}-\varepsilon_{it})^{2}\right)=O\left(N^{-1}\right)$. Also we have

$$E\left|\frac{1}{N}\sum_{i=1}^{N}\left(y_{is}-\tau(\rho)^{\frac{1}{2}}\rho^{s}\varepsilon_{i0}\right)\varepsilon_{it}\right|^{2}=\frac{1}{N^{2}}\sum_{i=1}^{N}E\left(\left(y_{is}-\tau(\rho)^{\frac{1}{2}}\rho^{s}\varepsilon_{i0}\right)^{2}\right)E\left(\varepsilon_{it}^{2}\right)=\frac{1}{N}O(1) \tag{151}$$

The first equality sign results from $\left(y_{is}-\tau(\rho)^{\frac{1}{2}}\rho^{s}\varepsilon_{i0}\right)$ and ε_{it} being independent with means zero since $s<t$ such that $\left(y_{is}-\tau(\rho)^{\frac{1}{2}}\rho^{s}\varepsilon_{i0}\right)\varepsilon_{it}$ is independent across i with mean zero. The second equality sign follows by using that $E\left(\varepsilon_{it}^{2}\right)=O(1)$ and $E\left(\left(y_{is}-\tau(\rho)^{\frac{1}{2}}\rho^{s}\varepsilon_{i0}\right)^{2}\right)=O(1)$. Altogether, this shows that $\frac{1}{N}\sum_{i=1}^{N}\left(y_{is}-\tau(\rho)^{\frac{1}{2}}\rho^{s}\varepsilon_{i0}\right)\Delta y_{it}\xrightarrow{P}0$ as $N\to\infty$ such that (141) is proved. To show the result in (142) we note that $Q_i'\,(\Delta y_{i,-1}-R_{1i}-R_{2i})$ has elements of the form $\tau(\rho)^{\frac{1}{2}}\rho^{s}\varepsilon_{i0}\left(\Delta y_{it}-(\rho-1)\,\tau(\rho)^{\frac{1}{2}}\rho^{t}\varepsilon_{i0}-\varepsilon_{it}\right)$ for $t=1,...,T-1$ and $s=1,...,t-1$. We have

$$\begin{aligned} &E\left|\frac{1}{N}\sum_{i=1}^{N}\tau(\rho)^{\frac{1}{2}}\rho^{s}\varepsilon_{i0}\left(\Delta y_{it}-(\rho-1)\,\tau(\rho)^{\frac{1}{2}}\rho^{t}\varepsilon_{i0}-\varepsilon_{it}\right)\right|^{2} \\ =\ & \frac{1}{N^{2}}\sum_{i=1}^{N}\tau(\rho)\,\rho^{2s}E\left(\varepsilon_{i0}^{2}\right)E\left(\left(\Delta y_{it}-(\rho-1)\,\tau(\rho)^{\frac{1}{2}}\rho^{t}\varepsilon_{i0}-\varepsilon_{it}\right)^{2}\right)=\frac{1}{N}O\left(N^{-1}\right) \end{aligned} \tag{152}$$

The first equality sign results from ε_{i0} and $\Delta y_{it}-(\rho-1)\,\tau(\rho)^{\frac{1}{2}}\rho^{t}\varepsilon_{i0}-\varepsilon_{it}$ being independent with means zero such that $\varepsilon_{i0}\left(\Delta y_{it}-(\rho-1)\,\tau(\rho)^{\frac{1}{2}}\rho^{t}\varepsilon_{i0}-\varepsilon_{it}\right)$ is independent across i with mean zero. The second equality sign follows by using that $\tau(\rho)=O(N)$, $\rho^{2s}E\left(\varepsilon_{i0}^{2}\right)=O(1)$ and

$E\left(\left(\Delta y_{it}-(\rho-1)\,\tau(\rho)^{\frac{1}{2}}\rho^{t}\varepsilon_{i0}-\varepsilon_{it}\right)^{2}\right)=O\left(N^{-1}\right)$. This shows that $E\left|\frac{1}{N}\sum_{i=1}^{N}Q_i'\left(\Delta y_{i,-1}-R_{1i}-R_{2i}\right)\right|^{2}\to 0$ as $N\to\infty$ which proves (142). Altogether, we have obtained the desired limits and the result in (a) is proved.

(b) We prove the result by showing that

$$\frac{1}{N^2}\sum_{i=1}^{N}Q_i'R_{2i}R_{2i}'Q_i\xrightarrow{P}\frac{1}{2\tilde{c}}\sigma_{4\varepsilon}\Gamma\quad\text{as } N\to\infty \tag{153}$$

$$\frac{1}{N^2}\sum_{i=1}^{N}\left(Z_{i2}-Q_i\right)'R_{2i}R_{2i}'Z_{i2}\xrightarrow{P}0\quad\text{as } N\to\infty \tag{154}$$

$$\frac{1}{N^2}\sum_{i=1}^{N}\left(Z_{i2}-Q_i\right)'R_{2i}R_{2i}'Q_i\xrightarrow{P}0\quad\text{as } N\to\infty \tag{155}$$

$$\frac{1}{N^2}\sum_{i=1}^{N}Z_{i2}'\left(\Delta y_{i,-1}-R_{2i}\right)R_{2i}'Z_{i2}\xrightarrow{P}0\quad\text{as } N\to\infty \tag{156}$$

$$\frac{1}{N^2}\sum_{i=1}^{N}Z_{i2}'\left(\Delta y_{i,-1}-R_{2i}\right)\Delta y_{i,-1}Z_{i2}\xrightarrow{P}0\quad\text{as } N\to\infty \tag{157}$$

The result in (153) follows directly by Markov's Law of Large Numbers and the remaining results are proved by using the same arguments as in the proof of Lemma 4. $\square$

Proof of Proposition 5:

According to the results in Lemma 6 have that when $\rho=1-c/\sqrt{N}$ for $c\geq 0$ and Assumption $2\,(i)$ is satisfied then

$$J_{II}\xrightarrow{w}\chi_m^2\left(q'\Sigma_{22}^{-1}q\right)\quad\text{as } N\to\infty \tag{158}$$

Using the definition in (119) we have

$$P_{11}^{-1}=\frac{1}{\sigma_\alpha^2\sigma_{2\varepsilon}+\tau\sigma_{4\varepsilon}} \tag{159}$$

$$P_{kk}^{-1}=\begin{bmatrix}\frac{\sigma_\alpha^2\sigma_{2\varepsilon}+(\tau+1)\sigma_{4\varepsilon}}{\sigma_{4\varepsilon}(\sigma_\alpha^2\sigma_{2\varepsilon}+\tau\sigma_{4\varepsilon})} & -\frac{1}{\sigma_{4\varepsilon}} & 0 & \cdots & 0\\ -\frac{1}{\sigma_{4\varepsilon}} & \frac{2}{\sigma_{4\varepsilon}} & \ddots & \ddots & \vdots\\ 0 & -\frac{1}{\sigma_{4\varepsilon}} & \frac{2}{\sigma_{4\varepsilon}} & \ddots & 0\\ \vdots & \ddots & \ddots & \ddots & -\frac{1}{\sigma_{4\varepsilon}}\\ 0 & \cdots & 0 & -\frac{1}{\sigma_{4\varepsilon}} & \frac{1}{\sigma_{4\varepsilon}}\end{bmatrix}\quad\text{for } k=2,...,T-1 \tag{160}$$

This implies that

$$\begin{aligned}q'\Sigma_{22}^{-1}q &= \sum_{k=1}^{T-1}q_k'P_{kk}^{-1}q_k=\sum_{i=1}^{T-1}\frac{(k-1)\,\sigma_{2\varepsilon}\left(\sigma_\alpha^2+\tau\sigma_{4\varepsilon}/\sigma_{2\varepsilon}\right)+\tau^2\sigma_{4\varepsilon}}{\sigma_{4\varepsilon}/\sigma_{2\varepsilon}\left(\sigma_\alpha^2+\tau\sigma_{4\varepsilon}/\sigma_{2\varepsilon}\right)}\\ &= \frac{\sigma_{2\varepsilon}^2}{\sigma_{4\varepsilon}}\left(\frac{(T-1)\,(T-2)}{2}+(T-1)\,\frac{\tau^2}{\sigma_\alpha^2\sigma_{2\varepsilon}/\sigma_{4\varepsilon}+\tau}\right)\end{aligned} \tag{161}$$

This proves the first part of the proposition.

The second part of the proposition follows by the results in Lemma 7 and by noting that the Moore-Penrose inverse of Γ denoted Γ^+ is given by

$$\Gamma^+ = \begin{bmatrix} \Gamma_1^+ & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \Gamma_{T-1}^+ \end{bmatrix} \tag{162}$$

where Γ_k^+ for $k = 1, ..., T-1$ is a symmetric $k \times k$ matrix with all elements equal to $1/k^2$. From above we have the following results

$$\frac{1}{N}\sum_{i=1}^{N} Z_{i2}'\Delta y_{i,-1} - \frac{1}{N}\sum_{i=1}^{N} Q_i'\left(R_{1i} + R_{2i}\right) \xrightarrow{P} 0 \quad \text{as } N \to \infty$$

$$\frac{1}{N^2}\sum_{i=1}^{N} Z_{i2}'\Delta y_{i,-1}\Delta y_{i,-1}' Z_{i2} - \frac{1}{N^2}\sum_{i=1}^{N} Q_i'\left(R_{1i} + R_{2i}\right)\left(R_{1i} + R_{2i}\right)' Q_i \xrightarrow{P} 0 \quad \text{as } N \to \infty$$

$$\left(\sum_{i=1}^{N} Q_i'\left(R_{1i} + R_{2i}\right)\left(R_{1i} + R_{2i}\right) Q_i\right)^{-\frac{1}{2}} \sum_{i=1}^{N}\left(Q_i'\left(R_{1i} + R_{2i}\right) + \frac{1}{2}\sigma_{2\varepsilon}\iota_m\right) \xrightarrow{w} N\left(0, I_m\right) \quad \text{as } N \to \infty$$

This proves that

$$J_{II} \xrightarrow{w} \chi_m^2\left(\tilde{c}\frac{\sigma_{2\varepsilon}^2}{\sigma_{4\varepsilon}}\frac{1}{2}\iota_m'\Gamma^+\iota_m\right) \quad \text{as } N \to \infty \tag{163}$$

and the result follows by using that $\iota_m'\Gamma^+\iota_m = T - 1$. $\square$

B Additional simulation results

This appendix contains the results from the simulation studies in Section 4 with $\sigma_\alpha^2 = 100$.

Table 5: Simulation results for statistics based on the Arellano-Bover type moment conditions under mean stationarity with $\tau(\rho) = 1$ and $\sigma_\alpha^2 = 100$

ρ	$T+1$	N	Mean $\hat{\rho}_I$	Std. $\hat{\rho}_I$	$P(t_I < q_1)$ *	$P(\bar{t}_I < q_1)$ *
0.900	5	100	0.9443	0.0804	0.2676 (0.7895)	0.3280 (0.7895)
0.900	5	250	0.9196	0.0587	0.6634 (0.9871)	0.8152 (0.9871)
0.900	5	500	0.9097	0.0399	0.9472 (0.9999)	0.9886 (0.9999)
0.900	5	1000	0.9046	0.0283	0.9990 (1.0000)	1.0000 (1.0000)
0.900	10	100	0.9587	0.0272	0.8436 (1.0000)	0.7066 (1.0000)
0.900	10	250	0.9331	0.0236	0.9936 (1.0000)	0.9954 (1.0000)
0.900	10	500	0.9184	0.0193	0.9998 (1.0000)	1.0000 (1.0000)
0.900	10	1000	0.9096	0.0146	1.0000 (1.0000)	1.0000 (1.0000)
0.950	5	100	0.9681	0.0670	0.1280 (0.3372)	0.1636 (0.3372)
0.950	5	250	0.9578	0.0443	0.3034 (0.6147)	0.4168 (0.6147)
0.950	5	500	0.9538	0.0285	0.6580 (0.8630)	0.7288 (0.8630)
0.950	5	1000	0.9519	0.0189	0.9464 (0.9871)	0.9596 (0.9871)
0.950	10	100	0.9760	0.0200	0.6440 (0.9123)	0.3776 (0.9123)
0.950	10	250	0.9643	0.0153	0.9362 (0.9990)	0.9164 (0.9990)
0.950	10	500	0.9580	0.0119	0.9994 (1.0000)	0.9996 (1.0000)
0.950	10	1000	0.9543	0.0087	1.0000 (1.0000)	1.0000 (1.0000)
0.975	5	100	0.9814	0.0653	0.0682 (0.1509)	0.1004 (0.1509)
0.975	5	250	0.9783	0.0393	0.1282 (0.2493)	0.1956 (0.2493)
0.975	5	500	0.9768	0.0232	0.2544 (0.3914)	0.3258 (0.3914)
0.975	5	1000	0.9759	0.0155	0.5108 (0.6147)	0.5428 (0.6147)
0.975	10	100	0.9869	0.0172	0.3874 (0.4424)	0.1674 (0.4424)
0.975	10	250	0.9815	0.0123	0.5954 (0.7663)	0.5016 (0.7663)
0.975	10	500	0.9787	0.0091	0.8732 (0.9563)	0.8542 (0.9563)
0.975	10	1000	0.9771	0.0064	0.9942 (0.9990)	0.9942 (0.9990)
0.990	5	100	0.9900	0.0636	0.0462 (0.0808)	0.0740 (0.0808)
0.990	5	250	0.9908	0.0371	0.0634 (0.1043)	0.1022 (0.1043)
0.990	5	500	0.9908	0.0208	0.0962 (0.1363)	0.1354 (0.1363)
0.990	5	1000	0.9904	0.0142	0.1624 (0.1921)	0.1948 (0.1921)
0.990	10	100	0.9943	0.0161	0.2196 (0.1480)	0.0828 (0.1480)
0.990	10	250	0.9925	0.0112	0.2264 (0.2432)	0.1658 (0.2432)
0.990	10	500	0.9915	0.0080	0.3188 (0.3809)	0.2900 (0.3809)
0.990	10	1000	0.9910	0.0055	0.5292 (0.5997)	0.5154 (0.5997)
1.000	5	100	0.9959	0.0612	0.0340 (0.0500)	0.0566 (0.0500)
1.000	5	250	0.9992	0.0364	0.0332 (0.0500)	0.0660 (0.0500)
1.000	5	500	1.0001	0.0199	0.0366 (0.0500)	0.0590 (0.0500)
1.000	5	1000	1.0001	0.0136	0.0418 (0.0500)	0.0572 (0.0500)
1.000	10	100	0.9995	0.0157	0.1360 (0.0500)	0.0474 (0.0500)
1.000	10	250	1.0000	0.0107	0.0802 (0.0500)	0.0552 (0.0500)
1.000	10	500	1.0002	0.0076	0.0580 (0.0500)	0.0528 (0.0500)
1.000	10	1000	1.0003	0.0052	0.0474 (0.0500)	0.0442 (0.0500)

* q_1 is the 5%-quantile of the standard normal distribution
The numbers in column 6-7 are the empirical rejection probabilities and the local power

Table 6: Simulation results for statistics based on the Arellano-Bover type moment conditions under covariance stationarity with $\tau(\rho) = 1/(1-\rho^2)$ and $\sigma_\alpha^2 = 100$

ρ	$T+1$	N	Mean $\hat{\rho}_I$	Std. $\hat{\rho}_I$	$P(t_I < q_1)$ *	$P(\bar{t}_I < q_1)$ *
0.900	5	100	0.9662	0.0945	0.1054 (0.3372)	0.2070 (0.6563)
0.900	5	250	0.9455	0.0866	0.2096 (0.6147)	0.5168 (0.8674)
0.900	5	500	0.9286	0.0769	0.4602 (0.8630)	0.8202 (0.9723)
0.900	5	1000	0.9136	0.0524	0.8192 (0.9871)	0.9762 (0.9989)
0.900	10	100	0.9764	0.0268	0.5366 (0.9123)	0.3892 (0.9853)
0.900	10	250	0.9563	0.0284	0.8134 (0.9990)	0.8642 (1.0000)
0.900	10	500	0.9377	0.0254	0.9766 (1.0000)	0.9894 (1.0000)
0.900	10	1000	0.9216	0.0207	1.0000 (1.0000)	1.0000 (1.0000)
0.950	5	100	0.9783	0.0838	0.0628 (0.1509)	0.1412 (0.4168)
0.950	5	250	0.9682	0.0808	0.0838 (0.2493)	0.3072 (0.5580)
0.950	5	500	0.9624	0.0661	0.1626 (0.3914)	0.5068 (0.7078)
0.950	5	1000	0.9557	0.0433	0.3692 (0.6147)	0.7548 (0.8674)
0.950	10	100	0.9859	0.0219	0.3542 (0.4424)	0.2238 (0.7510)
0.950	10	250	0.9771	0.0206	0.4834 (0.7663)	0.5868 (0.9393)
0.950	10	500	0.9689	0.0174	0.7708 (0.9563)	0.8934 (0.9943)
0.950	10	1000	0.9614	0.0138	0.9776 (0.9990)	0.9950 (1.0000)
0.975	5	100	0.9847	0.0724	0.0442 (0.0903)	0.1140 (0.3028)
0.975	5	250	0.9811	0.0697	0.0486 (0.1229)	0.2134 (0.3676)
0.975	5	500	0.9791	0.0677	0.0700 (0.1685)	0.3152 (0.4452)
0.975	5	1000	0.9771	0.0399	0.1282 (0.2493)	0.4644 (0.5580)
0.975	10	100	0.9909	0.0200	0.2662 (0.1854)	0.1470 (0.4711)
0.975	10	250	0.9874	0.0176	0.2640 (0.3231)	0.3418 (0.6419)
0.975	10	500	0.9841	0.0143	0.3810 (0.5128)	0.5878 (0.8036)
0.975	10	1000	0.9807	0.0111	0.6282 (0.7663)	0.8514 (0.9393)
0.990	5	100	0.9876	0.0751	0.0386 (0.0640)	0.0914 (0.2420)
0.990	5	250	0.9887	0.0616	0.0280 (0.0734)	0.1570 (0.2647)
0.990	5	500	0.9886	0.0602	0.0338 (0.0852)	0.2204 (0.2917)
0.990	5	1000	0.9898	0.0388	0.0514 (0.1043)	0.2848 (0.3317)
0.990	10	100	0.9946	0.0178	0.2082 (0.0893)	0.0942 (0.3007)
0.990	10	250	0.9938	0.0152	0.1466 (0.1209)	0.1920 (0.3639)
0.990	10	500	0.9931	0.0128	0.1498 (0.1650)	0.3022 (0.4397)
0.990	10	1000	0.9922	0.0098	0.2050 (0.2432)	0.4406 (0.5502)

* q_1 is the 5%-quantile of the standard normal distribution
The numbers in column 6-7 are the empirical rejection probabilities and the local power

Table 7: Simulation results for statistics based on the Arellano-Bond type moment conditions under mean stationarity with $\tau(\rho) = 1$ and $\sigma_\alpha^2 = 100$

ρ	$T+1$	N	Mean $(\hat{\rho}_{II} - \rho)$	Std. $(\hat{\rho}_{II} - \rho)$	$P(J_{II} > q_1)$*
0.900	5	100	-0.7092	0.4207	0.1500 (0.2076)
0.900	5	250	-0.4886	0.3743	0.4076 (0.5044)
0.900	5	500	-0.3057	0.2978	0.7388 (0.8471)
0.900	5	1000	-0.1783	0.2204	0.9784 (0.9942)
0.900	10	100	-0.6860	0.1537	0.4166 (0.8325)
0.900	10	250	-0.4672	0.1230	0.9910 (0.9998)
0.900	10	500	-0.3068	0.0944	1.0000 (1.0000)
0.900	10	1000	-0.1815	0.0668	1.0000 (1.0000)
0.950	5	100	-0.9021	0.4366	0.0744 (0.0821)
0.950	5	250	-0.7850	0.4439	0.1300 (0.1402)
0.950	5	500	-0.6319	0.4133	0.2320 (0.2559)
0.950	5	1000	-0.4556	0.3549	0.4648 (0.5044)
0.950	10	100	-0.8724	0.1685	0.1032 (0.2037)
0.950	10	250	-0.7326	0.1587	0.4728 (0.5583)
0.950	10	500	-0.5801	0.1422	0.8862 (0.9269)
0.950	10	1000	-0.4084	0.1133	0.9988 (0.9998)
0.975	5	100	-0.9739	0.4301	0.0588 (0.0575)
0.975	5	250	-0.9342	0.4468	0.0710 (0.0694)
0.975	5	500	-0.8661	0.4393	0.0886 (0.0910)
0.975	5	1000	-0.7710	0.4383	0.1410 (0.1402)
0.975	10	100	-0.9602	0.1683	0.0520 (0.0777)
0.975	10	250	-0.9062	0.1673	0.1374 (0.1330)
0.975	10	500	-0.8304	0.1671	0.2668 (0.2578)
0.975	10	1000	-0.7098	0.1578	0.5512 (0.5583)
0.990	5	100	-0.9983	0.4254	0.0522 (0.0512)
0.990	5	250	-0.9908	0.4368	0.0558 (0.0529)
0.990	5	500	-0.9699	0.4293	0.0588 (0.0559)
0.990	5	1000	-0.9513	0.4379	0.0636 (0.0621)
0.990	10	100	-0.9932	0.1661	0.0418 (0.0539)
0.990	10	250	-0.9831	0.1651	0.0726 (0.0602)
0.990	10	500	-0.9672	0.1665	0.0860 (0.0715)
0.990	10	1000	-0.9350	0.1665	0.1090 (0.0978)
1.000	5	100	-1.0041	0.4244	0.0520 (0.0500)
1.000	5	250	-1.0059	0.4369	0.0508 (0.0500)
1.000	5	500	-0.9951	0.4236	0.0492 (0.0500)
1.000	5	1000	-1.0019	0.4354	0.0546 (0.0500)
1.000	10	100	-1.0005	0.1653	0.0400 (0.0500)
1.000	10	250	-1.0015	0.1644	0.0606 (0.0500)
1.000	10	500	-1.0021	0.1648	0.0634 (0.0500)
1.000	10	1000	-0.9998	0.1631	0.0614 (0.0500)

* q_1 is the 95%-quantile of the χ^2-distribution with $\frac{1}{2}T(T-1)$ degrees of freedom
The numbers in column 6 are the empirical rejection probabilities and the local power

Table 8: Simulation results for statistics based on the Arellano-Bond type moment conditions under covariance stationarity with $\tau(\rho) = 1/\left(1-\rho^2\right)$ and $\sigma_\alpha^2 = 100$

ρ	$T+1$	N	Mean $(\hat{\rho}_{II} - \rho)$	Std. $(\hat{\rho}_{II} - \rho)$	$P(J_{II} > q_1)$*
0.900	5	100	-0.7809	0.4342	0.1126 (0.8429)
0.900	5	250	-0.6085	0.4018	0.2588 (0.9991)
0.900	5	500	-0.4259	0.3391	0.4890 (1.0000)
0.900	5	1000	-0.2702	0.2553	0.8360 (1.0000)
0.900	10	100	-0.7963	0.1630	0.1552 (0.9620)
0.900	10	250	-0.6333	0.1483	0.6928 (1.0000)
0.900	10	500	-0.4739	0.1277	0.9818 (1.0000)
0.900	10	1000	-0.3134	0.0966	1.0000 (1.0000)
0.950	5	100	-0.8705	0.4412	0.0816 (0.4998)
0.950	5	250	-0.7428	0.4302	0.1672 (0.9230)
0.950	5	500	-0.5811	0.3849	0.2946 (0.9991)
0.950	5	1000	-0.4055	0.3139	0.5858 (1.0000)
0.950	10	100	-0.8995	0.1673	0.0714 (0.6361)
0.950	10	250	-0.8023	0.1638	0.2766 (0.9915)
0.950	10	500	-0.6818	0.1566	0.5988 (1.0000)
0.950	10	1000	-0.5212	0.1326	0.9414 (1.0000)
0.975	5	100	-0.9064	0.4411	0.0720 (0.2535)
0.975	5	250	-0.8006	0.4376	0.1376 (0.6104)
0.975	5	500	-0.6576	0.4026	0.2322 (0.9230)
0.975	5	1000	-0.4808	0.3422	0.4532 (0.9991)
0.975	10	100	-0.9386	0.1676	0.0568 (0.2980)
0.975	10	250	-0.8720	0.1671	0.1626 (0.7694)
0.975	10	500	-0.7812	0.1658	0.3348 (0.9915)
0.975	10	1000	-0.6430	0.1502	0.6964 (1.0000)
0.990	5	100	-0.9334	0.4361	0.0630 (0.1189)
0.990	5	250	-0.8483	0.4406	0.1078 (0.2535)
0.990	5	500	-0.7261	0.4203	0.1784 (0.4998)
0.990	5	1000	-0.5584	0.3734	0.3434 (0.8429)
0.990	10	100	-0.9606	0.1672	0.0484 (0.1238)
0.990	10	250	-0.9125	0.1678	0.1196 (0.2980)
0.990	10	500	-0.8426	0.1695	0.2142 (0.6361)
0.990	10	1000	-0.7272	0.1609	0.4578 (0.9620)

* q_1 is the 95%-quantile of the χ^2-distribution with $\frac{1}{2}T(T-1)$ degrees of freedom
The numbers in column 6 are the empirical rejection probabilities and the local power

References

Anderson, T.W. and C. Hsiao, 1981, Estimation of dynamic models with error components, Journal of the American Statistical Association 76, 598-606.

Arellano, M., 2003, Panel data econometrics, (Oxford University Press, Oxford).

Arellano, M. and S. Bond, 1991, Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, Review of Economic Studies 58, 277-297.

Arellano, M. and O. Bover, 1995, Another look at the instrumental variable estimation of error-components models, Journal of Econometrics 68, 29-51.

Arellano, M., L.P. Hansen and E. Sentana, 1999, Underidentification?, Unpublished manuscript.

Baltagi, B.H., 1995, Econometric analysis of panel data, (Wiley, New York).

Baltagi, B.H., 2000, Advances in Econometrics vol. 15, Nonstationary panels, panel cointegration, and dynamic panels, (Elsevier, New York).

Baltagi, B.H. and C. Kao, 2000, Nonstationary panels, cointegration in panels and dynamic panels: A survey, in: Baltagi, B.H. (Ed), Advances in Econometrics vol. 15, 7-52.

Banerjee, A., 1999, Panel data unit roots and cointegration: an overview, Oxford Bulletin of Economics and Statistics, Special Issue, 607-629.

Blundell, R. and S. Bond, 1998, Initial conditions and moment restrictions in a dynamic panel data models, Journal of Econometrics 87, 115-143.

Blundell, R., S. Bond and F. Windmeijer, 2000, Estimation in dynamic panel data models: Improving on the performance of the standard GMM estimator, in: Baltagi, B.H. (Ed), Advances in Econometrics vol. 15.

Bond, S., C. Nauges and F. Windmeijer, 2002, Unit roots and identification in autoregressive panel data models: a comparison of alternative tests, Unpublished manuscript.

Breitung, J., 1997, Testing for unit roots in panel data using a GMM approach, Statistical Papers 38, 253-269.

Breitung, J. and W. Meyer, 1994, Testing for unit roots in panel data: Are wages on different Bargaining levels cointegrated?, Applied Economics 26, 353-361.

Hansen, L.P., 1982; Large sample properties of generalized method of moment estimators, Econometrica 50, 1029-1054.

Harris, R.D.F. and E. Tzavalis, 1999, Inference for unit roots in dynamic panels where the time dimension is fixed, Journal of Econometrics 91, 201-226.

Holtz-Eakin, D., W. Newey and H. Rosen, 1988, Estimating vector autoregressions with panel data, Econometrica 56, 1371-1395.

Kruiniger, H., 2002, GMM estimation of dynamic panel data models with persistent data, Queen Mary, University of London, Working Paper.

Madsen, E., 2003, Unit root inference in panel data models where the time-series dimension is fixed: A comparison of different tests, Ph.D. Thesis, University of Copenhagen.

Staiger, D. and J.H. Stock, 1997, Instrumental variables regression with weak instruments, Econometrica 65, 557-586.

White, H., 2001, Asymptotic Theory for Econometricians, Revised Edition, (Academic Press).